APPLICATION OF SMALL MOLECULE RECEPTORS TO THE ANALYSIS OF POST-TRANSLATIONAL MODIFICATIONS USING FLUORESCENT INDICATOR DISPLACEMENT ASSAYS.

Brendan Carroll Peacor

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry (Organic) in the College of Arts and Sciences

Chapel Hill
2015

Approved by:

Marcey Waters

Jeffrey Johnson

David Lawrence

Kevin Weeks

Stephen Frye

ABSTRACT

Brendan Carroll Peacor: Application of Small Molecule Receptors to the Analysis of Post-translational Modifications Using Fluorescent Indicator Displacement Assays
(Under the direction of Marcey Waters)

This dissertation focuses on the application of small molecule receptors to a fluorescent assay for the study of histone post-translational modifications, both in real-time enzyme reactions and endpoint characterization of analytes. In the first section, dynamic combinatorial chemistry was used to generate a series of **A_2X** receptors that varied the functionality of the **X** monomer. This allowed us to systematically study the contribution of pocket depth and electrostatic interactions on binding methylated lysine. We discovered that changing the location of a carboxylate increased affinity to $K(Me)_2$, presumably through a salt bridge, while an additional carboxylate increased affinity across the entire lysine series. Additionally, formation of a deeper binding pocket saw a selectivity increase for $K(Me)_3$ over the lower methylation states.

The remaining sections describe the application of the Waters lab suite of receptors to fluorescence indicator displacement assays (IDAs). In these assays, fluorescence signal is directly proportional to the competitive binding of a histone analyte. We applied a sensor system using the receptor **A_2N** and the fluorophore Lucigenin (**LCG**) to study the enzymatic dimethylation of histone H3 lysine 9 by the methyltransferase G9a. Optimization of the enzymatic buffer system established an effective methyltransferase reaction to short histone 3 peptide substrates. Applying these conditions to the fluorescent assay we are able to monitor

enzymatic activity, allowing future experiments to test enzyme response to neighboring modifications in the 'histone code'. This assay was also applied to the preliminary examination of the arginine methyltransferases, demonstrating the general applicability of the assay to the full range of enzymatic methylation reactions.

With the large number of receptors previously established, we sought to develop a general discriminatory assay capable of recognizing histone modifications beyond the designed scope of the sensor. By combining the fluorescent IDA signal for four different receptors, **A_2B**, **A_2D**, **A_2N**, and **A_2G**, we were able to accurately classify thirteen different histone peptides in a single output. Each peptide had multiple modifications, including arginine methylation and lysine methylation, as well as lysine methylation and threonine phosphorylation. The classification assay was able to distinguish both the degree of modification as well as the site of the specific modifications, all based on the slight perturbations neighboring residues make on binding affinity. This assay was also preliminarily applied to the sensing of complex enzymatic reactions by performing a mock kinase experiment, in which we were able to classify distinct 'time-point' of enzymatic phosphorylation on two separate substrates.

The final section focuses on the expansion of the target class of analytes for the combinatorial sensor array. While the previous study focused on modification-neighboring methylation events, here we describe the classification of the neutral modifications of arginine and lysine, reactions that abolish positive charge and weaken the affinity of the analytes to the receptor. Notably we are able to identify peptides based not just on how many charges are lost, but which specific residues are neutralized and where in the sequence the modification takes place. This opens the door to a large number of enzymatic reactions and histone analytes for which traditional methods of study are insufficient.

ACKNOWLEDGEMENTS

First I would like to thank Marcey Waters for your inspiration and mentorship over the last five years. You have allowed me to explore interesting science, even when it started us down an entirely new, and sometimes only slightly related path. You always answered my questions and had your own that challenged me to think of new solutions and ideas to try. I am so appreciative of the environment you have created in this lab, these five years have flown by in a supremely enjoyable and enriching way.

I would also like to thank my other committee members, Jeff Johnson, David Lawrence, Kevin Weeks, and Stephen Frye, as well as Mike Gagné who served on my orals committee. Your guidance and insight over the past several years has been extremely helpful and is very much appreciated.

Many thanks go out to the excellent collaborators I have been privileged to work with. Sam Pattenden and Jacqueline Norris-Drouin, your aid with G9a and in answering my general biology questions has been instrumental in getting these projects off the ground, thank you very much. Scott Rothbart, thank you for your help with the PRMT expression, and a special thanks to Brian Strahl for allowing me to use your lab space and radioactive assay.

To all Waters lab members past and present, thank you. Thank you for enjoying scientific discussions and making the workday something to look forward to. Nick, it has been a fantastic five years of lab work and out of lab fun, thanks for listening to my crazy ideas and being a great friend. Josh, thank you for mentoring me in lab when I started here and for working with me on

our receptor project. Masa, our conversations, both about science and the world in general have been supreme. The random weekend emails about comet landers and your general enthusiasm is awesome. Thank you Derek for all your help, both in lab and outside of it, looking forward to many more board game nights in the future. Amber and Effie, thanks for all your help, from running the peptide synthesizer to helping me figure out how to order proteins. To Kai, Jes, and Natalie, you have been wonderful to work with and get to know. Finally a special thanks to Chris Ramsay. Your work over the last two years was key to the success of the neutral histone code project.

Finally, I have many thanks for my family and friends who have supported me and made this journey possible. Scott, thanks for all the lunches and beers, it was great experiencing grad school with you. Kurt and Lindsey, Brian and Christopher, thank you for all your help and encouragement throughout these five years. To my best friends, Joannie and Ben, couldn't have done it without you. From the random weekends, to the heck of good fun new year's parties, you have kept my spirits high. Mom and Dad, your unyielding support and encouragement has been key my whole life. You taught me to be the person I am today, and if I spent the next 100 years repaying you it wouldn't be enough. My last, but critical thanks go to my wife, Allison. You have been there by my side this whole time, always supporting me, always putting up with my stressed out days and helping me forget them. I can't imagine being able to do this process without you, and can't wait to move on to our next adventures together!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Ac	Acetyl
AcCh	Acetylcholine
ACN	Acetonitrile
Ala, A	Alanine
AO	Acridine Orange
Arg, R	Arginine
Boc	tert-Butyloxycarbonyl
CB[#]	Cucurbit[#]uril
CH_3CN	Acetonitrile
ChIP	Chromatin Immunoprecipitation
Cit	Citrulline
CX#	Calix[#]arene
DA	Discriminate Analysis
DCC	Dynamic Combinatorial Chemistry
DCL	Dynamic Combinatorial Library
DMF	N,N-Dimethylformamide
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
Fmoc	Fluorenylmethoxycarbonyl
Gln, Q	Glutamine
Glu, E	Glutamic Acid
Gly, G	Glycine

HAT	Histone Acetyltransferase
HBTU	N,N,N′,N′-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HDAC	Histone Deacetylase
HP1	Heterochromatin Protein 1
HPLC	High Performance Liquid Chromatography
HOBt	Hydroxybenzotriazole
IDA	Indicator Displacement Assay
ITC	Isothermal Titration Calorimetry
K_d	Dissociation Constant
K(Me)	Monomethyllysine
$K(Me)_2$	Dimethyllysine
$K(Me)_3$	Trimethyllysine
LC-MS	Liquid Chromatography - Mass Spectrometry
LCG	Lucigenin
LDA	Linear Discriminate Analysis
Leu,L	Leucine
Lys/K/	Lysine
m/z	Mass to Charge Ratio
MS	Mass Spectrometry
NH_4OAc	Ammonium Acetate
NMR	Nuclear Magnetic Resonance
PAD	Protein Arginine Deiminase
Phe, F	Phenylalanine

Phos	Phosphorylation
PKMT	Protein Lysine Methyltransferase
PRMT	Protein Arginine Methyltransferase
PTM	Post-translational Modification
R(Me)	Monomethylarginine
$R(Me)_{2a}$	Asymmetric Dimethylarginine
$R(Me)_{2s}$	Symmetric Dimethylarginine
RP-HPLC	Reverse Phase-High Performance Liquid Chromatography
SAH/AdoHey	S-adenoxyl-L-homocysteine
SAM/AdoMet	S-adenosyl-L-methionine
Ser, S	Serine
TFA	Trifluoroacetic Acid
Thr, T	Threonine
TIPS	Triisopropylsilane
Trp, W	Tryptophan
Trt	Trityl
Tyr, Y	Tyrosine

CHAPTER 1 Structure Function Study of A_2X

1.1 Background

1.1.1 Ordered Structure of DNA

The genetic material responsible for coding eukaryotic life is contained within DNA and comprises the blueprints for all functions necessary to an organism's survival. DNA in humans is contained in 23 unique pairs of chromosomes each with a different complement of genes responsible for a variety of phenotypes. In order to encode this many genes chromosomes are quite large, with chromosome 1 alone comprising over 200 million base pairs for 3000 genes. With approximately 6 million base pairs over all 46 chromosomes, a linear strand of DNA would take up approximately 2 meters of space in a cell nucleus only 10 μm across.[1] Because of this, DNA must take on higher order structure to compact enough to be encapsulated within the cell, yet must still encode for the variety of proteins and biological functions required. To do this, DNA is wrapped into the nucleosome, which condenses into chromatin and from there can coil tightly into the traditional form of the chromosome.

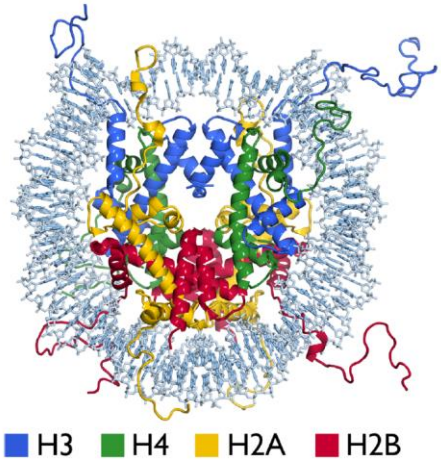


Figure 1.1 Crystal structure of a nucleosome. The DNA (in grey) wraps around 4 pairs of histone proteins (each colored as shown). From each of these proteins, an unstructured tail extends away from the complex.[2] Reprinted with permission from *Chem. Rev.* 2015, **115**, 2255-2273. Copyright 2015 American Chemical Society.

The nucleosome structure of DNA consists of 147 base pairs of DNA wrapped around 4 pairs of histone proteins, H2A, H2B, H3, and H4.[3] Each nucleosome is then connected like beads on a string through a linker region comprising of twenty to thirty base pairs wrapped around the linker histone protein H1 or H5.[2] The association of DNA to these histone proteins is most commonly thought of as a electrostatic attraction between the anionic DNA strand and the predominately cationic histone proteins. Importantly, the nucleosomes can be in one of two predominant forms, either heterochromatin or euchromatin, in which the DNA is either wrapped tightly around the histone and inaccessible for transcription or is loosely associated, resulting in activation of genes.[4]

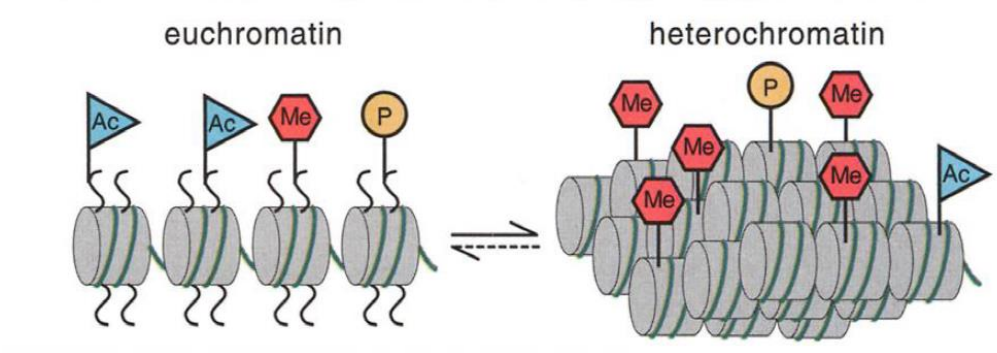


Figure 1.2 Representation of the two states of chromatin, either unwound and active (euchromatin) or tightly condensed and inaccessible (heterochromatin). Each state is host to a unique assembly of post-translational modifications.[5] From *Science*, 2001, **293**, 1074-1080. Reprinted with permission from AAAS.

Because each chromosome contains hundreds to thousands of genes, several mechanisms exist for switching between these active and inactive forms to preserve the compaction while ensuring no loss of function. A primary method for this switching utilizes the histone tails which extend outwards from the nucleosome complex, eight N-terminal tails, one from each histone, and two C-terminal tails, one from each H2A histone protein. [2] These tails are unstructured and are host to a variety of post-translational modifications (PTMs), covalent functional groups installed after the protein is synthesized.[6]

1.1.2 Post-translational Modifications

PTMs form a dynamic network of modifications, with various cellular machinery responsible for the ‘writing’, ‘erasing’, and ‘reading’ the marks dependent on the downstream event required.[7] There are a large variety of modifications possible, including methylation, acetylation, citrullination, and phosphorylation, among others, shown in Figure 1.3.

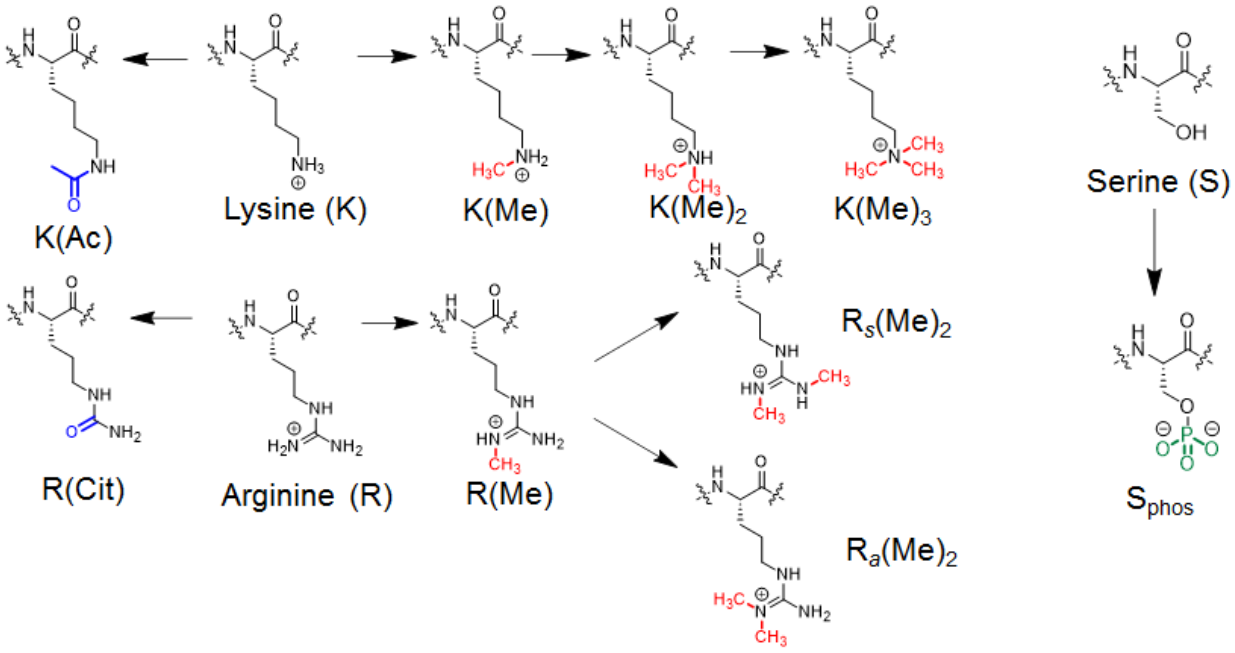


Figure 1.3 Post-translation modification of lysine, arginine, and serine. Each small chemical marker changes the recognition element of the amino acid side chain and is responsible for unique biological functions

In addition, larger molecules can be attached to the histone complex as PTMs, including sumoylation and ubiquitinilation, though these are outside of the scope of the current work.[8,9] Each of these modifications can influence the genetic landscape through direct interaction with DNA, as seen in phosphorylation of H4S47 or acetylation of the H3 tail or through the recruitment of 'reader' proteins that signal other events, such as trimethylation of H3K4 recruitment of the Chd1 remodeler protein.[10]

1.1.3 Lysine Methylation

1.1.3.1 Biological Significance

Lysine has three distinct methylation states, each installed by enzymes in the protein lysine methyltransferase family. Lysine methylation does not change the charge of the histone

tails, but rather is responsible for a diverse range of chromatin remodeling events. Critically, the downstream events rely not just on the degree of methylation, but also on the site of methylation.[11] This methylation can be responsible for transcriptional activation, in the case of methylated H3K4, H3K36, and H3K79, or transcriptional silencing in the case of methylated H3K9, H3K27, and H4K20.[6] Misregulation of the lysine methylation landscape has been implicated in many disease states. Expression of EZH2, responsible for the methylation of histone H3K27, has been linked to breast and prostate cancer, while loss of G9a, which dimethylates H3K9 has been shown to coincide with advantageous growth of cancer cells.[12] Lysine methylation is proposed to be largely responsible for the recruitment of other proteins to the histone-DNA complex, and these proteins can either operate on or further recruit other cellular factors responsible for biological consequences.

1.1.3.2 Reader Proteins for Methyl Lysine

Throughout the lysine reader protein family there are many different binding domains that are responsible for recognizing not just the degree of methylation but its sequence context as well.[13] As lysine is methylated, the hydrophobicity increases with a concurrent loss of hydrogen bonding capacity. Because of this, as well as the general hydrophobicity of the methylene units in the side chain, many lysine reader proteins rely on aromatic cages to recognize the modification. These cages have been described as either half cage or full cage motifs, comprised of Tyr, Phe, and Trp residues, with half cages generally including a residue capable of hydrogen bonding, as shown in Figure 1.4.[14]

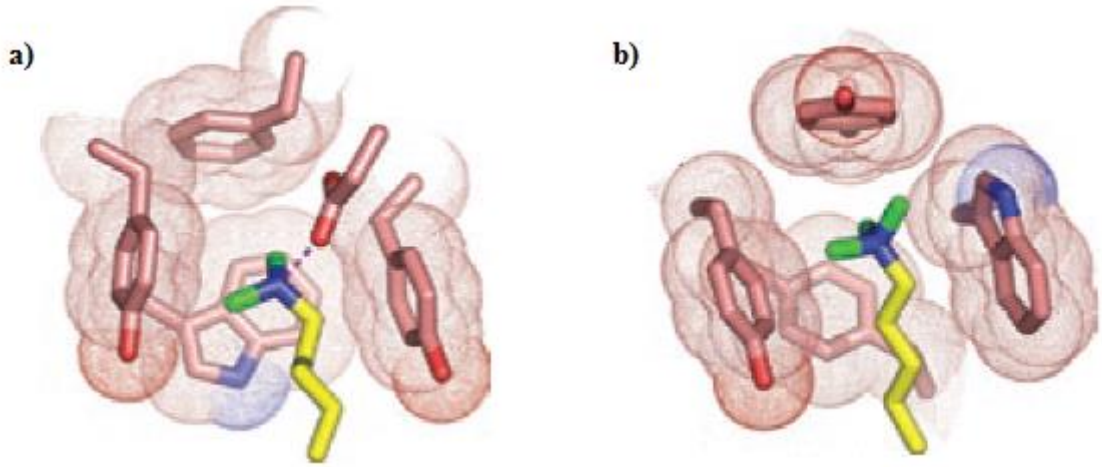


Figure 1.4 a) Dimethyl lysine bound to the pocket of 53BP1 b) Trimethyl lysine bound to the aromatic cage of BPTF.[13] Reprinted by permission from Macmillan Publishers Ltd: *Nat. Struct. Mol. Biol.*, 2007, **14**, 1025-1040. Copyright 2007.

As evidenced from Figure 1.4, 53BP1 favors $K(Me)_2$ binding through a variety of cation-π interactions as well as a hydrogen bonding interaction with the neighboring aspartic acid. $K(Me)_3$ on the other hand, because it lacks the capacity to hydrogen bond, is bound inside the pocket with four aromatic residues, creating an excellent hydrophobic pocket as well as cation-π contacts.[13] Work done in the Waters lab demonstrated that mutations of the binding pocket of reader proteins could significantly alter the binding properties and methylation state preference. The *Drosophila* heterochromatin binding protein, dHP1α can favorably bind to both trimethyl and dimethyl lysine 9 on the histone H3 tail using a half-cage motif of two tyrosines, a tryptophan, and a glutamic acid. Through mutation studies, they showed that the glutamic acid, while not participating in $K(Me)_3$ recognition, was important to binding $K(Me)_2$. By mutating the residue to glutamine, which abolishes the charge-charge interaction, the selectivity was improved for trimethyl lysine.[15] This suggests that a variety of residues and non-covalent interactions are critical for maintaining the binding affinities and selectivity necessary for biological function of these reader proteins.

In addition to the residues that make up the binding pocket, many reader proteins distinguish between the methylated forms of lysine by taking advantage of surface grooves or deep cavity binding sites, as shown in Figure 1.5.[16]

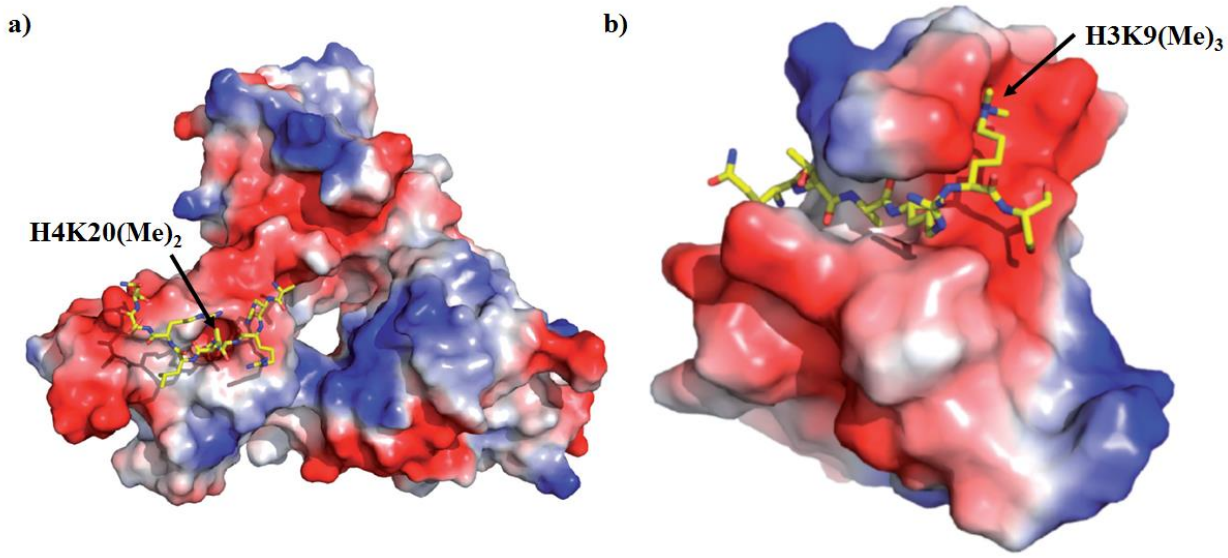


Figure 1.5 a) Histone H4K20(Me)$_2$ binding to L3MBTL1 through a deep cavity binding motif, where the dimethyl lysine side chain is buried into a pocket in the reader protein b) Histone H3K9(Me)$_3$ binding to HP1 through a surface groove motif, where the residue is surround by an aromatic cage but not buried.[16] Guo, Y, et al. Methylation-site-specific recognition of histones by the MBT repeat protein L3MBTL2. *Nucleic Acids Research*, 2009, **37**, *7*, 2204-2210 by permission of Oxford University Press.

The deep cavity binding of L3MBTL1 is selective for K(Me) and K(Me)$_2$, and L3MBTL1 binding is responsible for the compaction of chromatin.[17] The deep cavity motif imparts selectivity for lower methylation states because of its smaller size; K(Me)$_3$ is too bulky to thread in and interact with the aromatic cavity. Unmethylated lysine can thread into the cavity, but has weaker van der Waals' and cation-π interactions with the hydrophobic pocket, reducing overall affinity.[16,18] In contrast, the HP1 binding pocket is surface exposed, allowing the more bulky trimethyl lysine to bind, though selectivity is almost identical for K(Me)$_2$ and K(Me)$_3$, a general trend for surface binders due to the more promiscuous pocket.[13,15]

1.1.4 Non-covalent Binding of Methylated Lysine

1.1.4.1 Hydrophobic Effect

The classical hydrophobic effect describes the association of non-polar molecules in water. It postulates that in solution, water molecules around a hydrophobic surface, such as a hydrocarbon, become more 'ice-like', in that they engage in a higher number of hydrogen bonding interactions resulting in higher order and less favorable entropy. The association of two hydrophobic molecules reduces the surface area required for solvation, releasing the highly ordered water back into bulk solvent, and creating a favorable entropy term, as seen in Figure 1.6.[19]

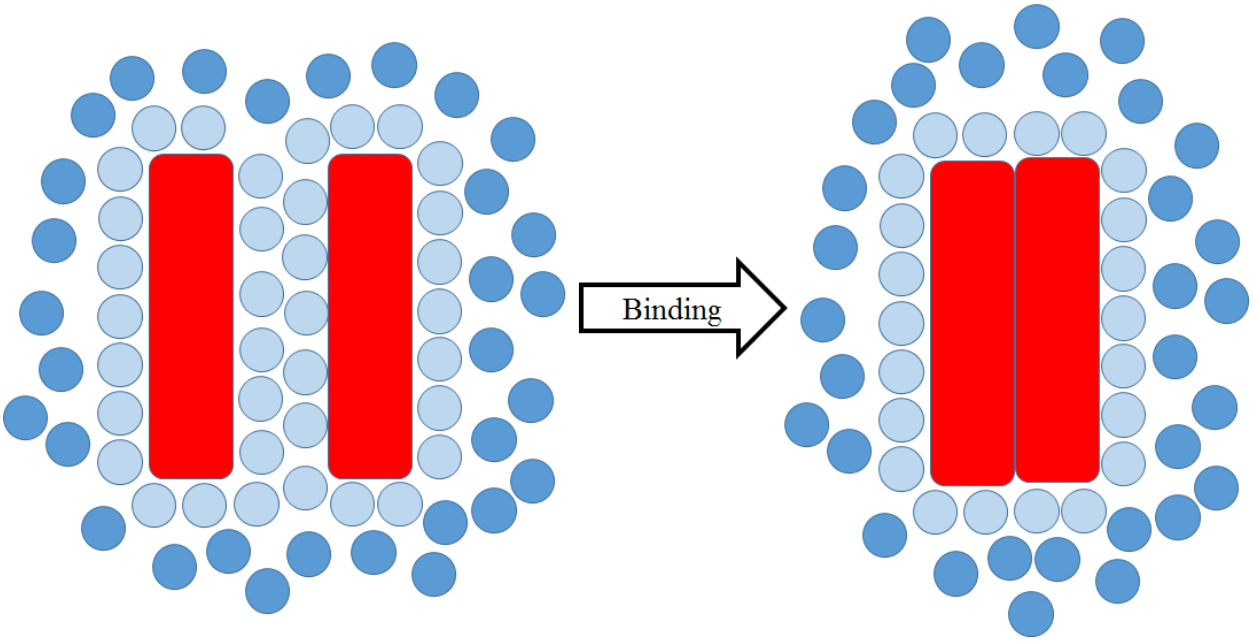


Figure 1.6 Classical hydrophobic effect. Ice-like water (Light blue) is released into bulk solvent (dark blue) upon association of two hydrophobic surfaces.

The classical hydrophobic effect has been used to explain a variety of events, including protein folding, micelle formation, and protein-ligand binding.[20,21] However, recent studies have shown a non-classical hydrophobic effect, in which enthalpic terms are the most favorable. This

enthalpy gain arose from the generation of other non-covalent interactions such as hydrogen bonding and cation-π interactions upon association, and these gains outweigh the disfavored entropy of restricted freedom.[22] Recently, work done by the Nau group has postulated another source of enthalpic gain upon complexation, the release of high energy water.[23] High energy water differs from highly ordered water in that high energy water cannot optimize its hydrogen bonding. This high energy water arises from the hydration of the inner cavities of molecules, of which every host species has a different size, and therefore different number of water molecules, as seen in Figure 1.7.

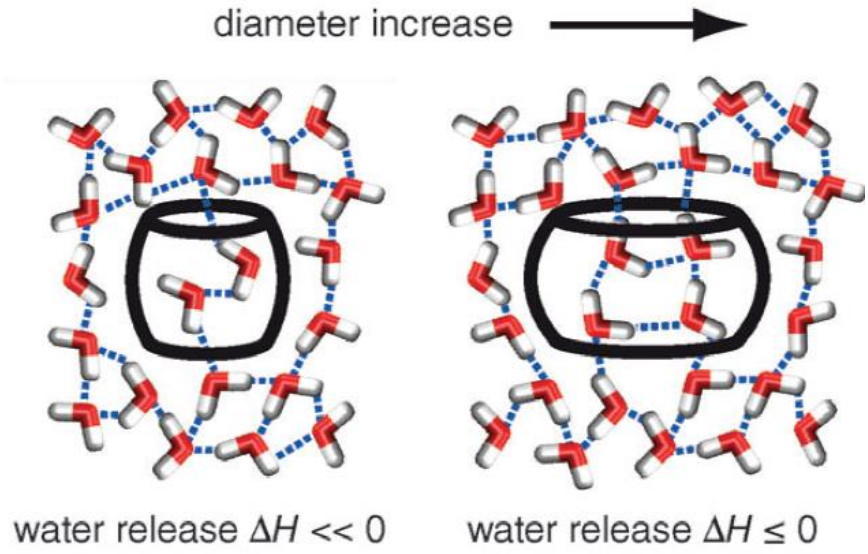


Figure 1.7 Diagram of the high-energy water inside two molecular containers. With increasing cavity size, more water molecules can fill the space, leading to less frustrated hydrogen bonding and lower energy complexes.[23] Reproduced with permission from Wiley: *Angew. Chem. Int. Ed.*, 2014, **53**, 11158-11171.

The study showed that in the cucurbituril family of receptors, cavity size was directly related to high energy water and therefore directly related to binding. The smaller cavities, CB6 and CB7, had strongly exothermic binding, due to high energy water release into bulk solvent,

whereas CB8 had minimal binding due to having enough inner cavity space to allow optimization of the water hydrogen bonding network.[24]

1.1.4.2 Cation-π Interactions

Another significant attractive force for biological molecules in aqueous environments is the cation-π interaction. Aromatic surfaces have a permanent quadrupole moment, such that there exist two regions above and below the aromatic ring with localized negative charge (Figure 1.8a).[25] Described in 1981 by Kebarle, then elaborated by Meot-Ner, the cation-π interaction is responsible for the stronger interaction between benzene and cations over the analogous water binding of said cations, as seen for tetramethylammonium in Figure 1.8b.[26–28] This motif is common in biological systems, with calculations showing that there is one significant cation-π interaction for every 77 residues within a given protein.[29] Additional studies have shown that while an electrostatic salt bridge decreases in affinity over 50-fold in water, the analogous cation-π interaction drops only 3-fold, allowing the interaction to remain viable in a wider variety of environments, both solvent-exposed or buried in a pocket.[30]

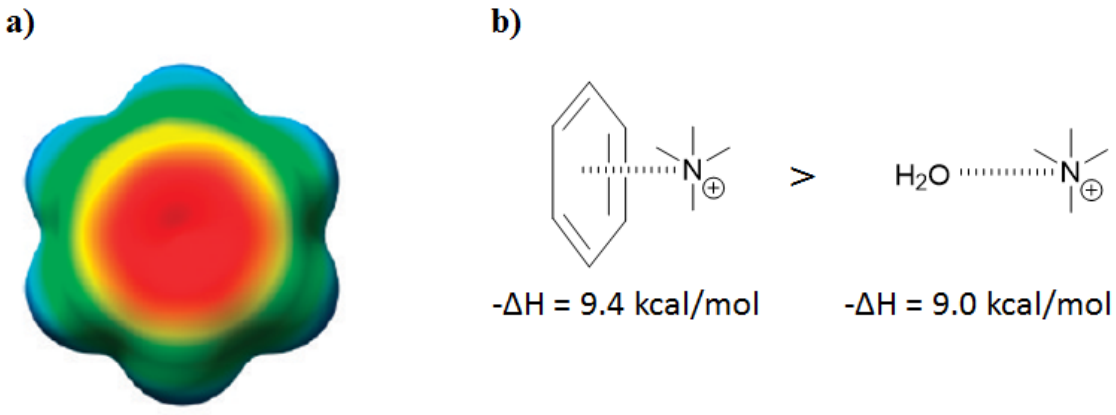


Figure 1.8 a) Electrostatic potential map of benzene[31] b) Association energy for tetramethylammonium interacting with benzene and water in the gas phase.[28] Reprinted with permission from *J. Chem. Theory Comput.*, 2009, **5**, 2301-2312

While critical and often referenced in biological function, the cation-π interaction is also prominent in small molecule recognition. In 1988, the Dougherty group designed a water-soluble cyclophane capable of binding to quaternary ammonium guest to explore the interaction, as shown in Figure 1.9.[32]

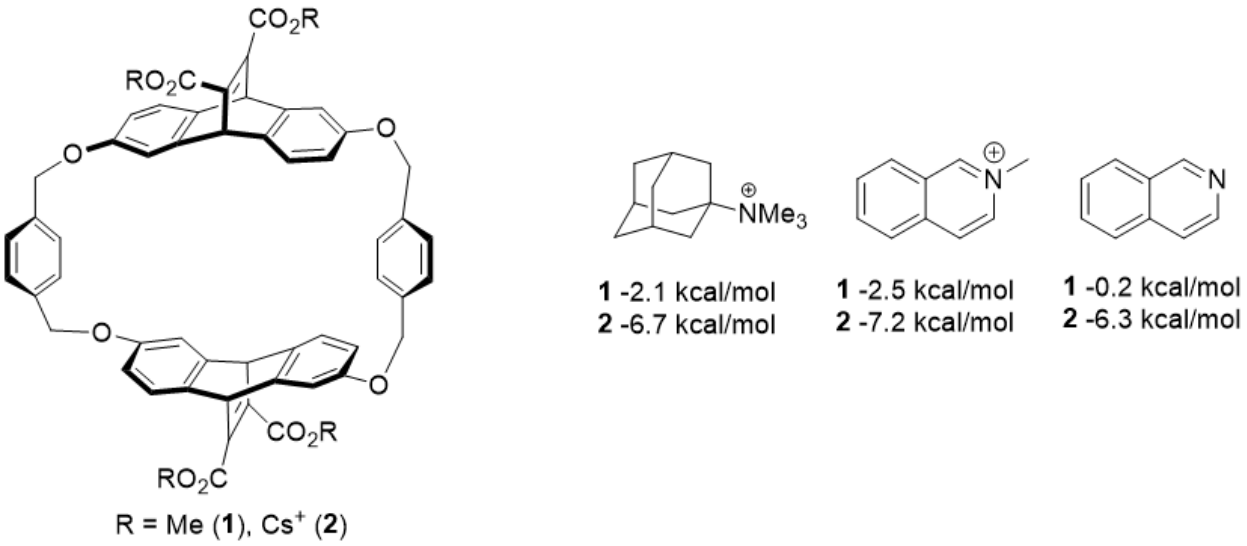


Figure 1.9 Cyclophane designed to complex quaternary ammonium species in $CDCl_3$ (**1**) or pD=9 buffer (**2**).[32]

In their work, they showed that the quaternary ammonium guests are able to tightly interact with the aromatic surface of their host (**2**), and still maintain their interaction in organic media (**1**), where solvent effects and ionic interactions to the charged carboxylates are nonexistent. Additionally, removal of the quaternary ammonium significantly reduced binding, to the point that the non-methylated guest did not bind in organic medium, ruling out electrostatics as the driving force of recognition. This result has been seen in other systems, such as analysis of the HP1 chromodomain binding to trimethyl lysine studied in the Waters lab. Upon mutation of $K(Me)_3$ to tert-butyl norleucine the neutral analog, the binding affinity of the native chromodomain drops 30 fold, from 10 μM to 310 μM, highlighting the importance of the cation-π effect in recognition.[33]

1.1.5 Small Molecule Receptors for Trimethyl Lysine

1.1.5.1 Cucurbiturils

In recent years much work has gone into the development and characterization of small-molecule receptors capable of binding to methylated lysine targets. In 2013, the Macartney lab studied the binding of methylated lysine and arginine to the supramolecular host cucurbituril.[34] Cucurbiturils (CB) have previously been shown to bind to a variety of cationic guests due to their hydrophobic cavity with dense negative charge at either rim. This allows binding to hydrophobic guests that bear positive charge, ideal for methylated lysine.[35] The Macartney group studied the interaction between CB6 and CB7 to the various methylated lysines in their amino acid form. They observed a 3500-fold selectivity for binding to $K(Me)_3$ vs the unmethylated lysine, with a K_d of 500 nM for $K(Me)_3$. NMR studies showed that the N-methyl groups experience upfield shifting similar to that of other cations, and neutralization of the amino acid carboxylate resulted in tighter binding, suggesting that binding was negatively influenced by the close proximity of the anionic residue to the carbonyl rim.[34]

1.1.5.2 Calixarenes

Recently the calixarene series of hosts has also been used to bind to trimethyl lysine in both binding studies and biological applications. The Hof group reported in 2010 that the p-sulfonatocalix[4]arene (CX4) host could differentiate between the methylated forms of lysine in both amino acid and peptide based forms.[36] They observed almost 20-fold selectivity for trimethyl lysine over the unmethylated form in the short peptide Ac-RKST-NH_2, with a K_d for $K(Me)_3$ of around 10 μM. The high affinity and selectivity of binding was proposed to come from a combination of interactions. The lysine residues can all experience cation-π interactions

with the CX4 cavity, but only the $K(Me)_3$ residue is capable of burying into the hydrophobic core of the receptor, as the lower methylation states are too well solvated. Hof then performed a redesign and optimization of the CX4 receptor through selective modification of a single aromatic ring to achieve higher binding to trimethyl lysine, as shown Figure 1.10.[37]

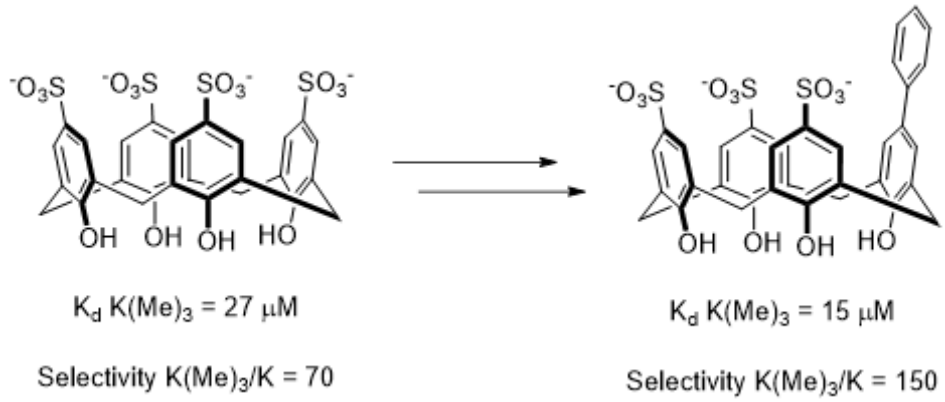


Figure 1.10 Modification of the CX4 receptor (left) into the trisulfonated host (right). Affinities shown based on NMR titrations of the $K(Me)_3$ and unmethylated lysine amino acids.

Using a multi-step synthesis the Hof group was able to furnish several derivatives of the trisulfonated calixarene, though only one displayed higher affinity binding and selectivity for trimethyl lysine. They hypothesized that the extra phenyl substituent was able to make increased contact with the side chain of lysine, increasing affinity, though additional substitution on this phenyl group significantly hampered binding. These calixarene receptors have been further applied to the disruption of protein-protein interactions between trimethylated lysine and several reader proteins, highlighting the utility of applying synthetic receptors to biological problems.[38,39]

1.1.6 Dynamic Combinatorial Chemistry

The synthesis of novel receptors can be quite challenging, requiring several steps of low yield followed by isolation and characterization of each new receptor to determine if

improvements were made to binding affinity and selectivity. Dynamic combinatorial chemistry (DCC) has arisen as a favorable alternative to this method of preparing larger macrocycles, allowing for both affinity screening and synthesis to occur in the same step. DCC is defined as combinatorial chemistry under thermodynamic control, where library members can exchange until the most stable, and therefore favorable, species is selected.[40] The inclusion of a guest, should it create a more thermodynamically favorable species, will shift the equilibrium of the library such that said species is amplified, as shown in Figure 1.11.[41]

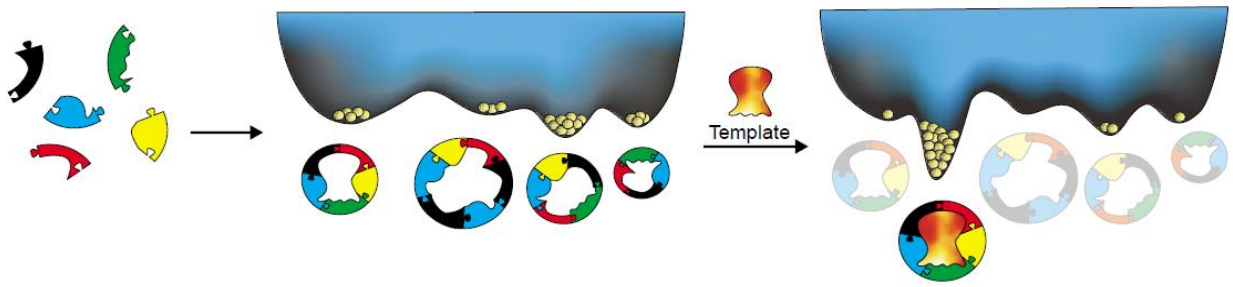


Figure 1.11 Dynamic combinatorial chemistry. Monomers in solution form stable species. The addition of a guest or template induces an equilibrium shift to favor the most stable binder.[41] Reprinted from *Curr. Opin. Chem. Biol.*, **6**, Otto, S., Furlan, R., Sanders, J., Recent Developments in Dynamic Combinatorial Chemistry. 321-327, Copyright 2002, with permission from Elsevier.

There have been a large number of reversible exchange reactions developed, including imine exchange, metathesis reactions, and thiol-thioester exchange to name a few.[42–45] Disulfide exchange is also used to generate macrocyclic species due to its ease of exchange and ability to perform at near physiological pH, as well as having an easily adapted synthesis of thiol functionalized building blocks.[46] The libraries are initiated at pH 7-9 to allow formation of the initial thiolate species which can perform a nucleophilic attack on the disulfide species present in the library to furnish a reversible reaction that is easily halted by lowering the pH. This utility has made disulfide exchange one of the most common DCC reactions.

1.1.6.1 Disulfide Receptors for Methylated Lysine

In 2008, Otto and Sanders applied disulfide exchange to the cyclophanes designed by Dougherty and described in section 1.1.4.2. They were able to split the cyclophane into several components that bore aromatic thiol functionalities with carboxylates for water solubility, and allowed these components to assemble in the presence of quaternary ammonium guests.[47]

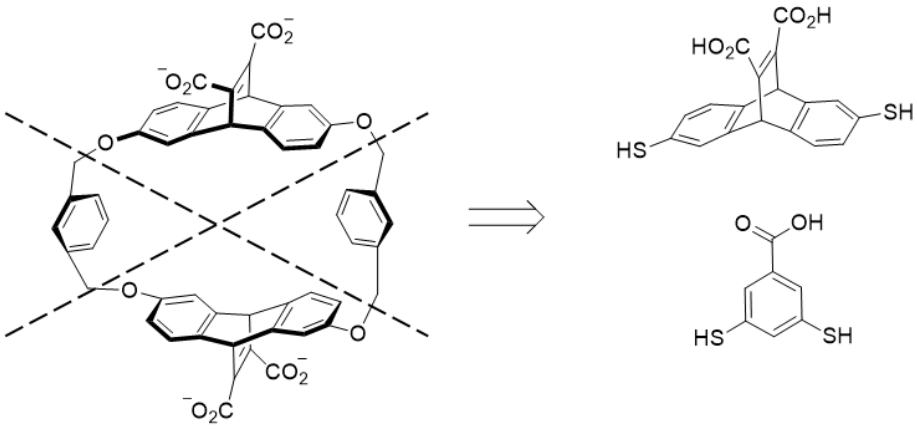


Figure 1.12 Retrosynthesis of Dougherty's cyclophane into its constituent dithiol building blocks for DCC.

Based on this work, the Waters group established the use of disulfide exchange to study methylated lysine receptors. They screened a variety of monomers that contained carboxylates for water solubility, aromatic cores for cation-π and hydrophobic effect interactions, and dithiols for reversible exchange.[48] Using DCC, they discovered the small molecule receptor **A_2B**, which bound to $K(Me)_3$ with native protein like affinity, and showed selectivity over the lower methylation states of lysine. The receptor was proposed to bind to trimethyl lysine through cation-π interactions, with selectivity arising from the increased cost of desolvation for the various lower methylated states.

In 2014, the Waters group followed up **A_2B** with a successful iterative monomer redesign of monomer **B** to furnish the $K(Me)_3$ receptor **A_2N (**Figure 1.13**)**.[49]

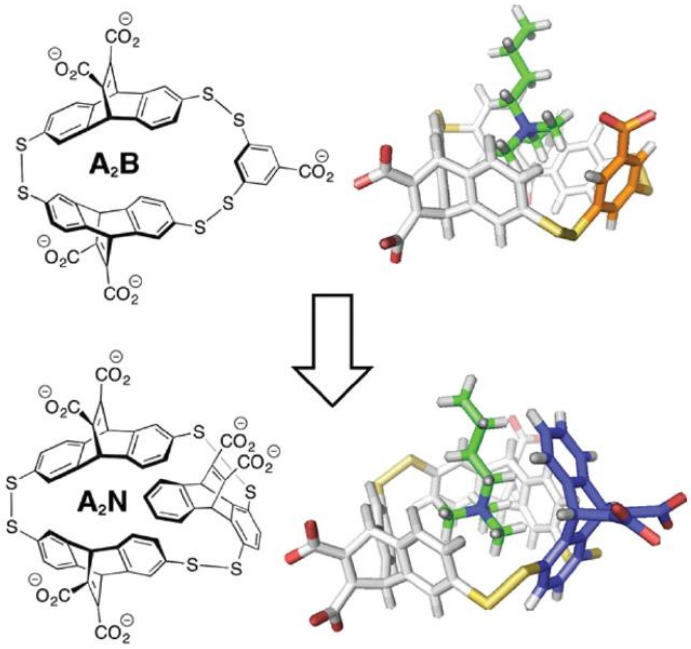


Figure 1.13. Redesign of the small molecule receptor **A2B** to **A2N** to deepen the binding cavity as well as increase cation-π interactions.[49] Reproduced from *Org. Biomol. Chem.*, 2014, *12*, 7059-7067 with the permission of The Royal Society of Chemistry.

This receptor bound to $K(Me)_3$ with high nanomolar affinity, with the extra aromatic surface providing a 1.3 kcal/mol increase in binding affinity, similar to that reported for other cation-π interactions. Additionally, the extra ring on the **N** monomer provided a more hydrophobic surface, increasing the favorable entropic gain upon binding to trimethylated lysine based on the classical hydrophobic effect. This elegant redesign of a single monomer enabled the facile screening and synthesis of a tighter binder, highlighting the utility of DCC in creating synthetic host systems.

1.1.7 Motivation

The work done to develop **A2B** and **A2N** demonstrated the synthetic ease of rapidly synthesizing novel macrocyclic species with precise control over their monomeric subunits. This

afforded a unique opportunity to study the contributions of a variety of non-covalent interactions in the molecular recognition of methylated lysine. By modifying the monomeric species, we can set up dynamic combinatorial libraries (DCLs) that allow for simultaneous monitoring of amplification as well as synthesis of host. This highlights the utility of DCC in structure function studies and receptor optimization, and allows us to draw comparisons between our receptors and the reader proteins that natively recognize this class of modification to further drive the discovery and understanding of chemical biology tools for PTMs.

1.2 Results

1.2.1 Dynamic Combinatorial Libraries

1.2.1.1 Monomer E

With each of our previously described receptors, we saw retention of the **A_2X** motif, suggesting that this was a highly favorable minimum energy for any library containing monomer **A**. One of the main advantages of DCC is that we could easily monitor libraries for selectivity through amplification, making it facile to prepare several libraries in tandem and perform a high-throughput monomer screen. For this study, we chose to prepare monomer **E** (Figure 1.19), which was reported in 2006 by the Otto group to bind to the poly amine spermine.[50] We prepared DCC libraries containing monomer **A** and **E** to examine if it would form the **A_2X** motif as well as libraries combining monomer **B** and **E** to test the possibility of forming tetramer structures like those previously reported. All libraries were equilibrated in the presence of the peptides Ac-$K(Me)_nGGY\text{-}NH_2$, which allowed us to examine effect of lysine methylation on the library members. The peptides were synthesized with a tyrosine tag to simplify HPLC purification as well as act as a concentration tag for accurate library concentrations.

The DCC libraries containing **B** and **E** had only two main peaks amplified, with masses corresponding to the $\mathbf{B_3}$ trimer and $\mathbf{B_4}$ tetramer, as seen in Figure 1.14.

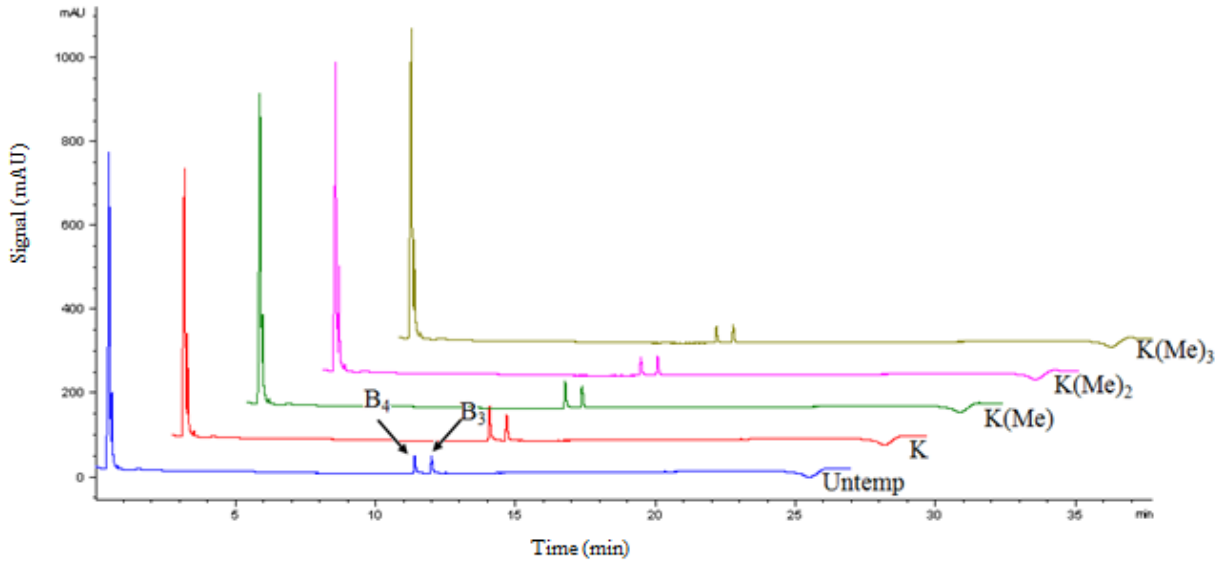


Figure 1.14 DCL containing monomers **E** and **B** (1 mM each) with peptide guest Ac-K(Me)$_n$GGY (2 mM) monitored at 254 nm.

The $\mathbf{B_3}$ and $\mathbf{B_4}$ macrocycles were favored in the unmethylated lysine library and amplification decreases as methylation state increases. Interestingly, we saw no incorporation of the **E** monomer, though the trace signal at 254nm is higher at 0.5 minutes for the later libraries, suggesting that the **E** containing macrocycles are not being retained on the column, though no masses corresponding to macrocyclic species could be distinguished.

As we expected, the library containing monomers **A** and **E** formed exclusively the $\mathbf{A_2E}$ macrocycle, as shown in Figure 1.15.

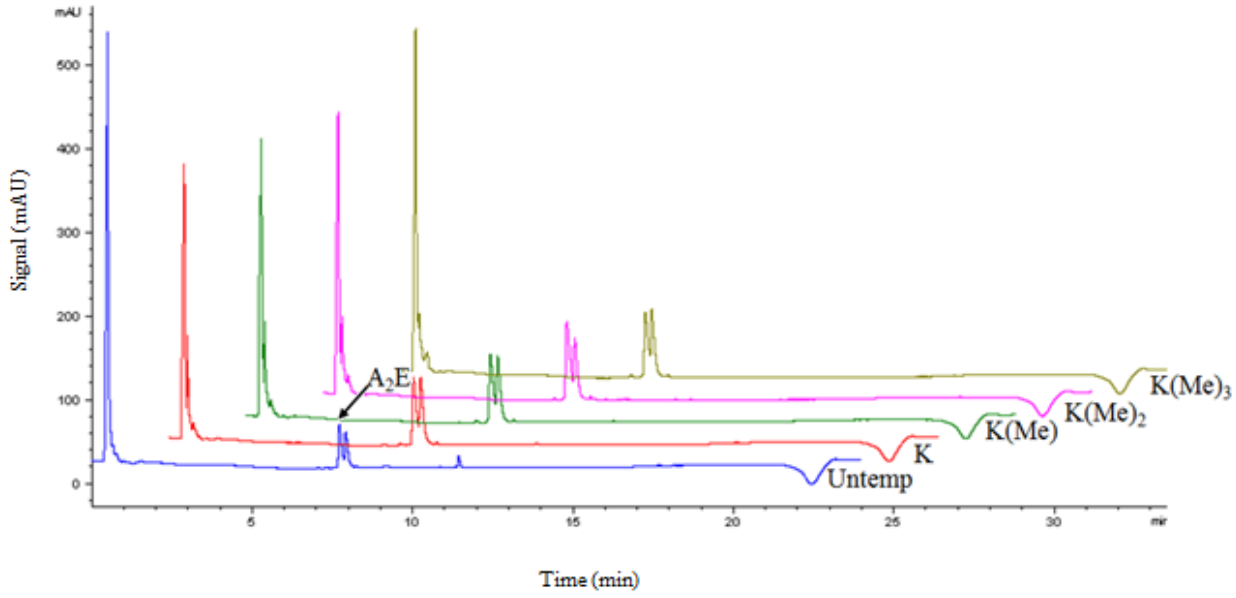


Figure 1.15 DCL containing monomers **E** and **A** (1 mM each) with peptide guest Ac-K(Me)$_n$GGY (2 mM) monitored at 254 nm.

Our initial libraries, prepared at millimolar concentrations of monomers and not in a buffered system, gave surprising amplification, suggesting that the dimethylated lysine amplified **A$_2$E** the most. However, upon analysis of the library using different conditions, we observed that the peaks corresponding to peptide in the K(Me)$_3$ library were drastically lower, suggesting that in the non-buffered system there was precipitation, skewing the amplification. Therefore, we re-ran the libraries using 50 mM borate buffer, pH 8.5 with lower concentration of monomer and specifically biased towards the formation of **A$_2$E**. These libraries displayed the anticipated selection for trimethyl lysine as the tightest binder, as shown below in Figure 1.16 and Figure 1.17.

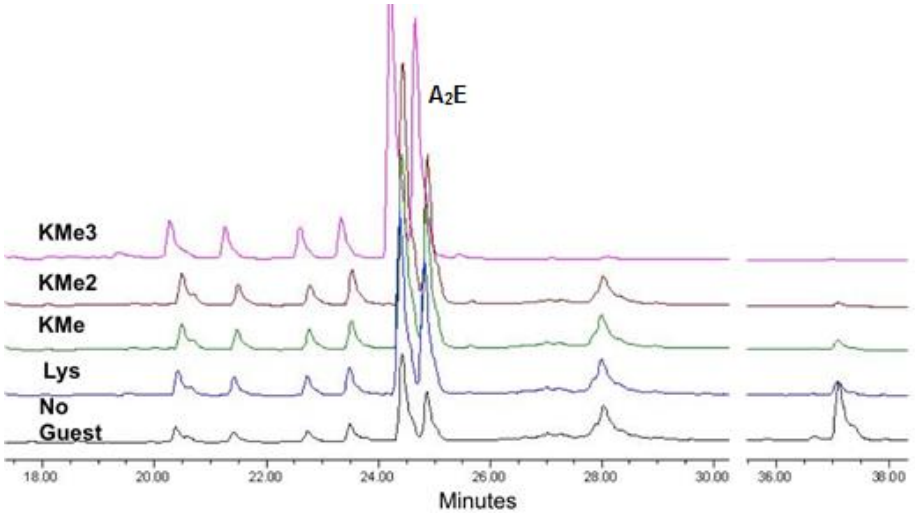


Figure 1.16 DCL containing monomers **A** (0.5 mM) and **E** (0.25 mM) and Ac-K(Me)$_n$GGY-NH_2 (0.75 mM), monitored at 280 nm.

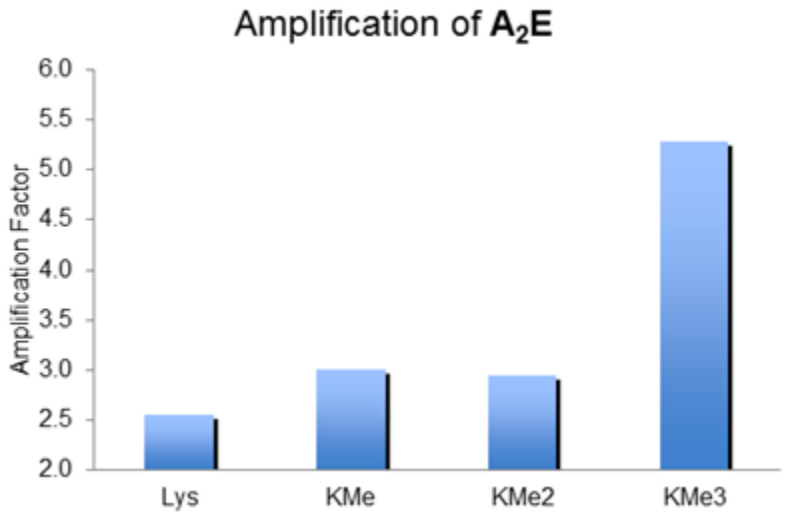


Figure 1.17 Amplification data for the low concentration DCC libraries biased toward the formation of **A$_2$E** with 0.5 mM **A**, 0.25 mM **E**, and 0.75 mM Ac-K(Me)$_n$GGY-NH_2, as compared to the untemplated library.

The receptor **A$_2$E** was prepared on a preparative scale upon combination of **A** and **E** (2 mM each) with the small molecule acetyl choline (8 mM) as a mimic of KMe3, followed by incubation for one week and purification using semi-preparative HPLC with ammonium acetate eluents. Upon isolation of the receptor as a mixture of isomers, the extinction coefficient was

determined by mass and binding characterization was performed using Isothermal Titration Calorimetry (ITC). Concurrently with this receptor, another member of the lab, Dr. Joshua Beaver, discovered several other monomers that amplified the **A_2X** motif. Because each library amplified the **A_2X** framework, we established a structure function study to characterize the contribution of modifying the single **X** residue.

1.2.1.2 Additional Monomers Characterized in This Study

The additional receptors used in this study were synthesized and characterized by Dr. Joshua Beaver, and full experimental details can be found in the resulting publication.[51] Figure 1.18 shows the structure of each monomer studied as part of the **A_2X** framework.

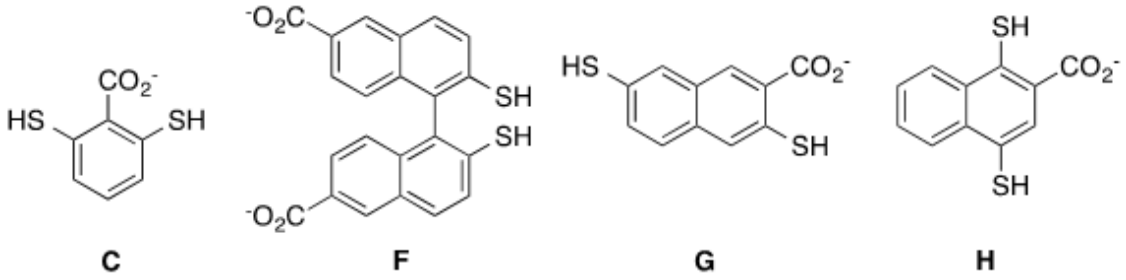


Figure 1.18 Monomers synthesized by Dr. Joshua Beaver for the structure function study of **A_2X**.

For each monomer, exploratory libraries were set up to test for amplification in the presence of methylated lysine. Monomer **C**, designed to test the effect of moving the carboxylate closer to the interior of the receptor pocket, saw amplification of the **A_2C** receptor. The amplification was the largest and similar between dimethyl and trimethyl lysine, suggesting that this receptor is capable of interacting strongly with both species (Figure 1.22). For monomer **G**, which was synthesized to examine a larger binding pocket, we again observed amplification of

the **A_2G** species, this time heavily favoring the trimethyl lysine species (Figure 1.23), suggesting a selective binder. For monomers **F** and **H**, we did not observe useful library amplification. Monomer **F** did not display any higher order macrocyclic species upon addition of lysine guests, and monomer **H**, while forming the **A_2H** species, displayed equally high amplification regardless of guest (Figure 1.24). We hypothesized that this monomer was capable of self-templating **A_2H** through edge-face interactions between the naphthyl moiety of **H** and the aromatic surface of the **A_2** cleft, therefore abolishing selectivity.[49]

1.2.2 Investigation of Receptors with Varied Electrostatic Interactions

We first investigated the binding of methyl lysine bearing peptides to the receptors **A_2C** and **A_2E**, which varied the electrostatic interactions present inside the methyl lysine binding pocket of the **A_2X** framework. We studied the binding of these receptors to peptide mimics of the histone H3 tail, Ac-WGGG-QTARK$(Me)_n$STG-NH_2 using all four methylation states of lysine. The WGGG tag was appended to the N-terminus to allow accurate concentration determination, and the control peptide Ac-WGGG-QTAGGSTG-NH_2 was also examined. This control peptide showed no binding to any of the studied receptors in isolation, though there could still be contributions to overall binding that are present only in combination with the methylated lysine interaction.

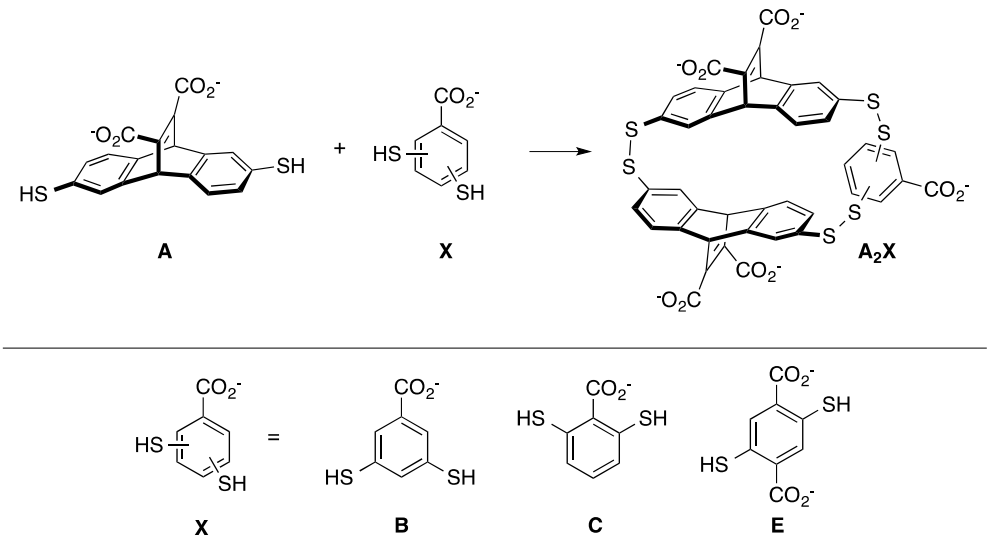


Figure 1.19 Monomers used to examine the contribution of electrostatic interactions in the **A2X** framework.

Table 1.1 Thermodynamic binding data obtained for binding of **A2B**, **A2C**, and **A2E** to Ac-WGGG-QTARK(Me)nSTG-NH2 as measured by ITC.[a]

Entry	Receptor[b]	Peptide	K_d [c] (μM)	Selectivity [d]	ΔG [c] (kcal/mol)
1 [e]	**A2B**	KMe3	2.6 ± 0.1	-	-7.63 ± 0.03
2[e]	**A2B**	KMe2	6.3 ± 0.3	2.4	-7.10 ± 0.07
3[e]	**A2B**	KMe	13.9 ± 0.1	5.4	-6.64 ± 0.01
4[e]	**A2B**	Lys	22 ± 1	8.3	-6.38 ± 0.02
5 [e]	**A2B**	R8GKME3	17.1 ± 0.1		-6.52 ± 0.01
6	**A2C**	KMe3	2.3 ± 0.1	-	-7.69 ± 0.02
7	**A2C**	KMe2	2.8 ± 0.2	1.2	-7.57 ± 0.04
8	**A2C**	KMe	13.8 ± 0.7	6.0	-6.63 ± 0.03
9	**A2C**	Lys	22 ± 1	9.6	-6.34 ± 0.03
10	**A2C**	R8GKME3	29 ± 3		-6.17 ± 0.05
11[f]	**A2E**	KMe3	0.191 ± 0.002	-	-9.16 ± 0.01
12[f]	**A2E**	KMe2	0.5 ± 0.1	2.6	-8.5 ± 0.1
13 [f]	**A2E**	KMe	1.6 ± 0.2	8.4	-7.92 ± 0.08
14 [f]	**A2E**	Lys	6.7 ± 0.1	35	-7.05 ± 0.01
15[f]	**A2E**	R8GKME3	2.7 ± 0.3		-7.59 ± 0.06

(a) All data determined by ITC, fit to one-site binding model; Condi-tions: 26 °C, in 10 mM sodium borate buffer, pH 8.5. (b) All receptors are mixtures of isomers except rac-A2B. (c) Errors are from averages of three trials, unless noted otherwise. (d) Selectivity is calculated as the fold difference in affinity for KMe3 over the designated methyla-tion state of the peptide in that row. (e) Data reported by Pinkin and Waters.21 (f) Average of two trials.

Binding experiments revealed that **A2C**, with the carboxylate moved relative to the original receptor **A2B** to place it *ortho* and between the two thiols, displayed no change in

affinity for $K(Me)_3$, K(Me), or unmethylated lysine. However, the binding affinity for $K(Me)_2$ increased by approximately 2-fold, essentially abolishing selectivity of the receptor for $K(Me)_2$ and $K(Me)_3$ in a similar fashion to shallow binding pocket reader proteins.[13] Molecular modeling of **A_2C** revealed a potential hydrogen bonding interaction within the pocket, as seen in Figure 1.20.

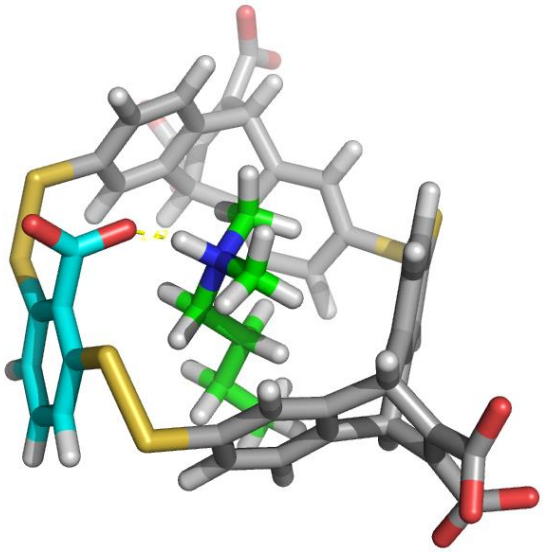


Figure 1.20 Molecular modeling of **A_2C** (**A** in grey, **C** in cyan) binding to butyldimethyl ammonium (green).

The modeling suggests that the methyl groups of lysine can associate with the aromatic walls of the **A_2** cleft through several cation-π effects, providing several favorable binding interactions. Interestingly, the positioning of the carboxylate *ortho* to the two thiols twists it out of the plane of the aromatic ring of **C**. This conformation is what directs the charge inside the binding pocket, becoming available for hydrogen bonding. This hydrogen bond does not influence $K(Me)_3$ binding, presumably because it can still fully participate in the interactions seen previously in **A_2B**, and the lower methylation states are likely well solvated to fit in the pocket, disfavoring binding even with the addition of the hydrogen bond.

Addition of the extra carboxylate in **A_2E** increased the binding affinity to trimethylated lysine by an order of magnitude compared to **A_2B** (Table 1.1, entries 1 and 11). Additionally, the binding to each form of methylated lysine was increased by 1.3-1.5 kcal/mol, though the selectivity across this series remained essentially unchanged. This result suggests that the extra charge provides a consistent increase in binding, irrespective of the degree of methylation, similar to work done by Dougherty in which increased receptor charge enhanced binding to well solvated guests through cooperative interaction.[52] However, the binding selectivity over unmethylated lysine increased dramatically, from 8 fold in the case of **A_2B** to 35 fold for **A_2E** due to a lesser overall increase in binding affinity for unmethylated lysine.

Previous work had shown that the neighboring charge can influence receptor binding to trimethyl lysine (Table 1.1, entry 5).[49] By varying the amino acid sequence and replacing the lysine-adjacent arginine with glycine (R8G), we can examine the specific binding of receptors to trimethyl lysine. For both **A_2C** and **A_2E** this resulted in a 13-fold drop in affinity corresponding to a loss of 1.5 kcal/mol of binding energy. In comparison, **A_2B** only drops 7 fold when arginine is swapped for glycine, suggesting that by repositioning the negative charge or increasing it, the receptor is more sensitive to the neighboring peptide sequence. However, these results do highlight the ability to fine-tune the binding properties of a synthetic receptor through addition or modification of the existing electrostatic contributions, though only a small difference in selectivity was observed.

1.2.3 Investigation of Receptors with Deeper Binding Pockets

Previous work in the Waters lab had seen the iterative re-design of **A_2B** into receptor **A_2N**, which bound to $K(Me)_3$ with high nanomolar affinity.[49] Interestingly, this affinity increase was thought to arise due to the much deeper binding pocket of the receptor. While **A_2B**, and now

A_2C and **A_2E,** had shallow binding pockets with electrostatic interactions close to the interior, **A_2N** cannot form electrostatic interactions to guests bound inside the pocket. Additionally, monomer **N** has an additional aromatic surface responsible for greater cation-π interactions with the methylated ammonium as well as CH-π interactions with the side chain methylenes (Figure 1.21). This deeper pocket allows trimethyl lysine to bury inside, while the lower methylation states require more significant desolvation because the inner cavity is deeper.

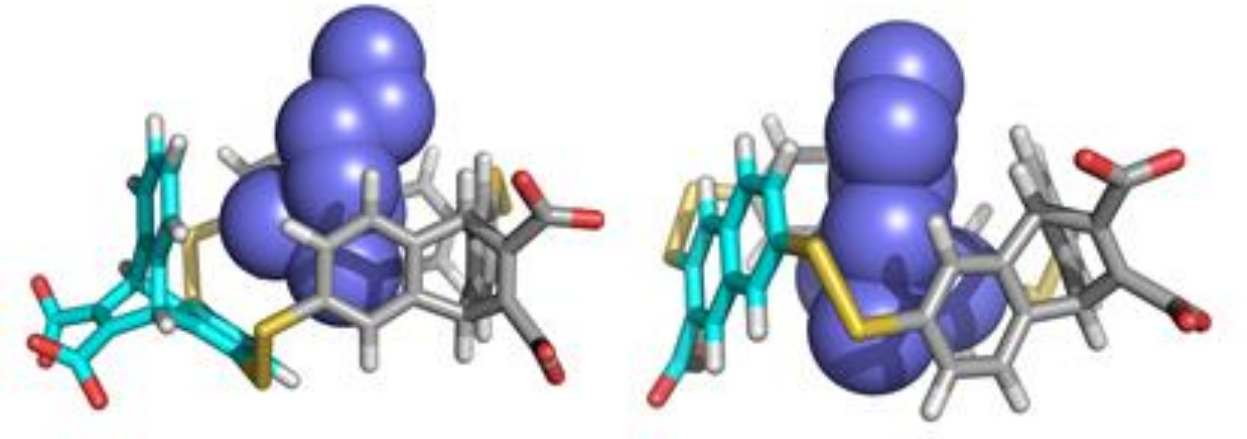


Figure 1.21 Molecular models of **A_2N** (left) and **A_2G** (right) bound to butyl-trimethyl ammonium (blue) as a model for $K(Me)_3$.

As previously stated, monomer **G** was investigated to study the contribution of a larger binding pocket relative to **A_2B**. However, modeling suggests that instead of a more flat, open pocket, **A_2G** adopts a twisted conformation, creating a deeper binding pocket, as shown in Figure 1.21.

Table 1.2 Thermodynamic binding data obtained for binding of **A_2N** and **A_2E** to Ac-WGGG-QTARK(Me)nSTG-NH2 as measured by ITC.[a]

Entry	Receptor[b]	Peptide	K_d [c] (μM)	Selectivity [d]	ΔG [c] (kcal/mol)
1 [e]	**A_2N**	KMe3	0.30 ± 0.04	-	-8.91 ± 0.07
2 [e]	**A_2N**	KMe2	4.1 ± 0.5	14	-7.36 ± 0.04
3 [e]	**A_2N**	KMe	40 ± 4	130	-6.01 ± 0.06
4 [e]	**A_2N**	Lys	10.5 ± 0.9	35	-6.80 ± 0.05
5 [e]	**A_2N**	R8GKMe3	1.3 ± 0.2		-8.05 ± 0.08
6 [f]	**A_2G**	KMe3	1.4 ± 0.1	-	-8.00 ± 0.05
7 [f,g]	**A_2G**	KMe2	13.2 ± 2.4	10	-6.6 ± 0.1
8 [f]	**A_2G**	KMe	15 ± 1	11	-6.57 ± 0.04
9 [f,h]	**A_2G**	Lys	>58	>40	< -5.8
10 [fg]	**A_2G**	R8GKMe3	5.4 ± 0.1		-7.19 ± 0.01

(a) All data determined by ITC, fit to one-site binding model; Conditions: 26 °C, in 10 mM sodium borate buffer, pH 8.5. (b) **A_2N** was measured as the *meso*-species, **A_2G** as a mixture of isomers. (c) Errors are from averages of three trials, unless noted otherwise. (d) Selectivity is calculated as the fold difference in affinity for KMe3 over the designated methylation state of the peptide in that row. (e) Data reported by Pinkin and Waters[21] (f) Average of two trials. (g) Error determined by propagation from curve fitting and averages. (h) These values are approximate because the c-value for these experiments was <1.

ITC revealed that **A_2G** bound to $K(Me)_3$ about two fold tighter than **A_2B** (K_d = 1.4 ± 0.1 μM). Additionally, it had good selectivity over the lower methylated species of lysine, comparable to **A_2N**, supporting the hypothesis of a deeper, more selective binding pocket. Binding for $K(Me)_2$ was similar to that of K(Me), suggesting that neither of the smaller guests fit well in the pocket, and the selectivity over unmethylated lysine was larger than 40-fold, the greatest observed so far, presumably due to the higher surface area of the pocket and increased desolvation cost of the guest.

Comparison of the R8G mutation for trimethyl lysine binding revealed that the arginine residue contributes 0.8 kcal/mol to peptide binding, which is smaller than its contribution to the shallower binding pockets of **A_2B**, **A_2C**, and **A_2E**. This suggests that the deeper binding pockets are more selective for the targeted residue, and that the shallow electrostatic binding pockets are more influenced by the neighboring peptide sequence.

1.3 Conclusion

In summary, we have used DCC to perform a structure function study on the **A_2X** macrocyclic framework to establish the contributions of several different non-covalent binding interactions. DCC allowed for both rapid affinity screening and facile synthesis of several different receptors, with varying electrostatic interactions and pocket depths. Interestingly, the fast screening methodology allowed for several monomers to be screened that did not synthesize useful, selective binders. This allowed us to move on quickly from the unproductive monomers without expending time in the full synthesis and characterization, a large advantage of DCC.

Of the monomers and receptors we did synthesize, we observed that the shallow binding pockets were influenced by not just the number of electrostatic interactions but also their position in the receptor. By adding an extra carboxylate in monomer **E**, we observed an increase in binding affinity across the series of methylated lysine, but by repositioning the carboxylate to *ortho* between the thiols we saw a significant increase in affinity for only the $K(Me)_2$ species, indicative of a hydrogen bond. This is significant because it shows that by careful manipulation of the functionality of monomers it would be possible to create selective receptors for the lower methylated forms of lysine.

We also explored the contribution of binding from a deeper binding pocket and observed that the selectivity of binding for $K(Me)_3$ increased. This is presumably due to the lower desolvation cost of $K(Me)_3$, allowing it to bury into the pocket. Interestingly the deep pocket also showed the least influence of neighboring charge on guest binding, suggesting that this motif could lead to a pan-selective receptor for trimethyl lysine.

These receptors reinforce many of the binding motifs seen in biological reader proteins for methylated lysine, despite the fact that they are many orders of magnitude smaller. In addition, the utility of DCC in rapidly designing, screening, and synthesizing receptors cannot be understated, fueling novel receptor design and broadening of targets to other methylated species and furthering the understanding of non-covalent binding interactions in water.

1.4 Experimental

Full synthesis and characterization, including ITC for receptors **A_2C**, **A_2G**, and **A_2H** can be found in the publication of this work.[51]

1.4.1 Peptide Synthesis

All peptide synthesis was performed on a Tetras Peptide Synthesizer using Peptides International CLEAR-Amide resin. Peptides were synthesized on a 0.06 mmol scale. All amino acids with functionality were protected during synthesis. Coupling reagents were HOBt/HBTU in DMF. For the dipeptides, the N-terminus was acylated with a solution of 5% acetic anhydride and 6% 2,6-lutidine in DMF. Cleavage was performed by hand with a cocktail of 95% TFA/2.5% triisopropylsilane/2.5% H_2O for 3 hours.

Methylated peptides were synthesized with either 2 equivalents of Fmoc-Lys(Boc)(Me)-OH purchased from BaChem or Fmoc-Lys(Me)2-OH•HCl purchased from Anaspec and coupled for 4 hours. The trimethyl lysine-containing peptides were synthesized by reacting the corresponding dimethylated peptides (0.6 mmol scale) prior to cleavage from the resin with MTBD (10.8 μL, 0.075 mmol) and methyl iodide (37.4 μL, 0.6 mmol) in DMF (5 mL) for 5 hours with bubbling N_2 in a peptide synthesis flask stoppered with a vented septum. After

washing the resin with DMF (3x), CH_2Cl_2 (3x), and drying, the peptide was cleaved and purified as normal.

Peptides were purified by semipreparative reverse-phase HPLC on a C18 column at a flow rate of 4 mL/min. Peptides were purified with a linear gradient of A and B (A: 95% H_2O, 5% CH_3CN with 0.1% TFA, B: 95% CH_3CN, 5% H_2O with 0.1% TFA) and elution was monitored at 214 nm. Once purified, peptides were lyophilized to powder and characterized by ESI-MS. Peptides used in binding studies were desalted and repurified by semipreparative reverse-phase HPLC with a C-18 column and buffered mobile phase. Peptides were purified using an optimized gradient of A and B (A: 100% H_2O, 10 mM NH_4OAc; B: 90% CH_3CN, 10% H_2O, 10 mM NH_4OAc). The ammonium salts were removed under reduced pressure for three to five days after the samples were dry.

1.4.2 DCC Libraries

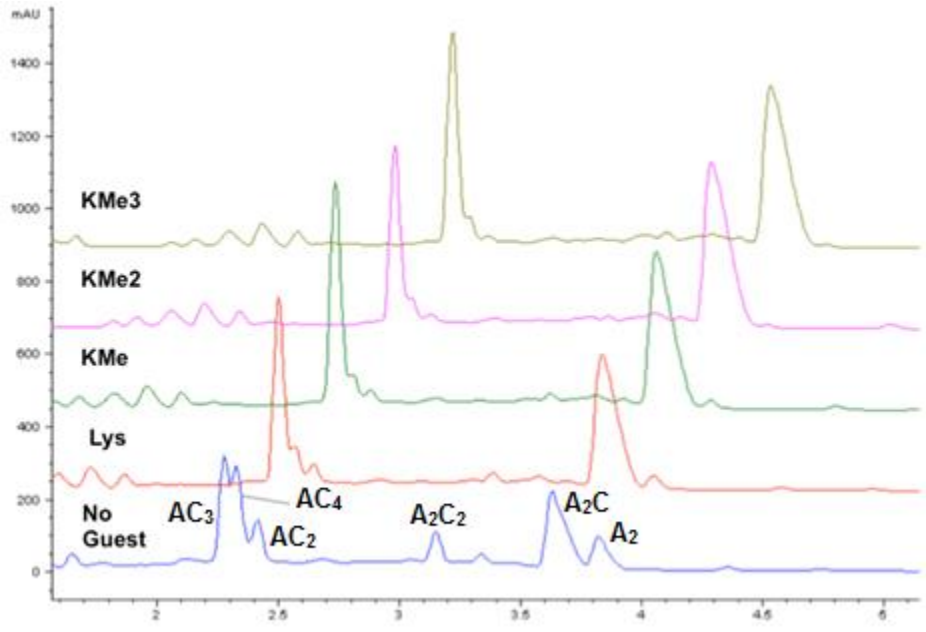


Figure 1.22 Overlaid HPLC traces at 254 nm of DCC libraries biased toward the formation of **A_2C** with monomers **A** (5 mM) and **C** (2.5 mM) and Ac-K$(Me)_n$G-NH_2 (7.5 mM).[5]

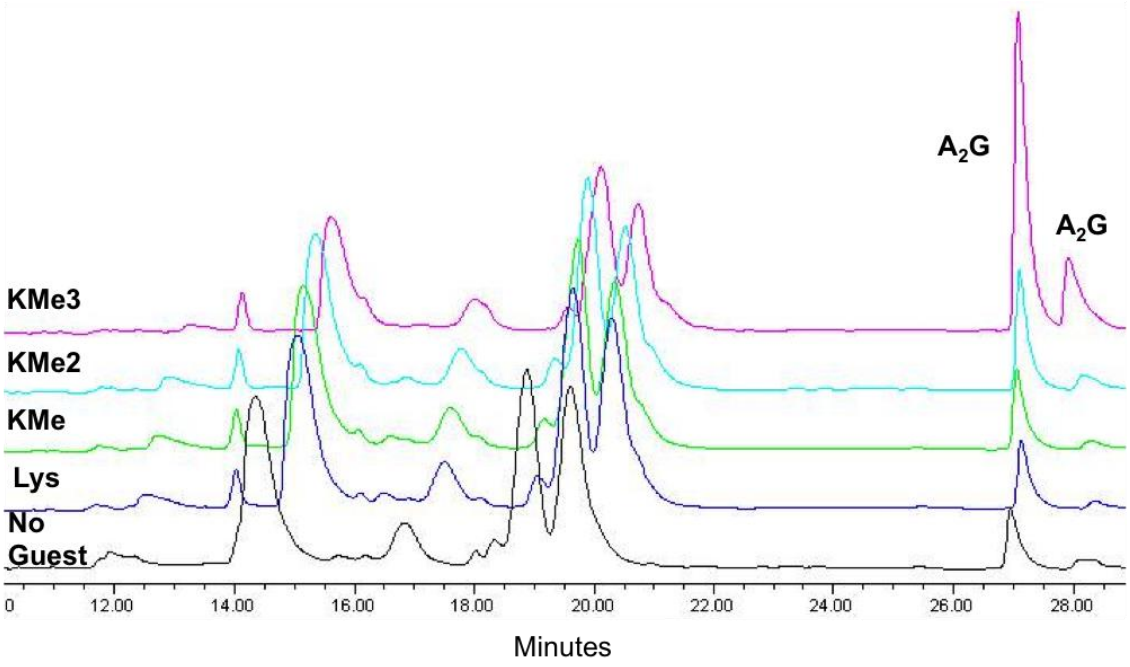


Figure 1.23 Overlaid HPLC traces at 254 nm of DCC libraries biased toward the formation of **A_2G** with monomers **A** (0.5 mM) and **G** (0.25 mM) and Ac-K(Me)$_n$GGY-NH_2 (0.75 mM).

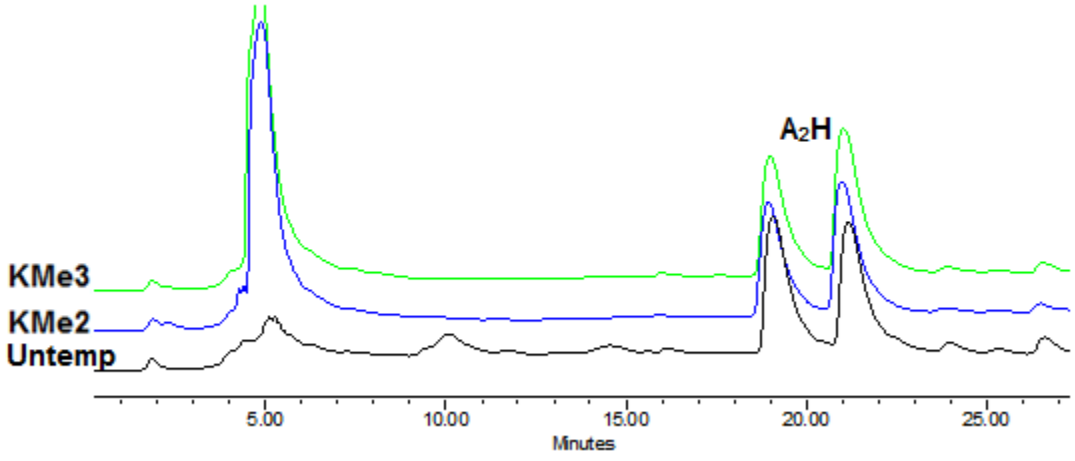


Figure 1.24 Overlaid HPLC traces at 254 nm of DCC libraries biased toward the formation of **A_2H** with monomers **A** (5 mM) and **H** (2.5 mM) and Ac-K(Me)$_n$GGY-NH_2 (7.5 mM).

1.4.3 Preparative Synthesis of A_2E

A_2E was synthesized using a preparative scale DCL, with monomer **A** (2.0 mM), monomer **E** (2.0 mM) and **Acetylcholine Chloride** (**AcCH**) (8.0 mM) in 50 mM sodium borate

buffer, pH = 8.5. **AcCH** was used because it is commercially available and amplified **A_2E** under preparative library conditions. After five days, the library was filtered and purified using semi-preparative HPLC (solvent A: 10mM NH_4OAc in H_2O; solvent B: 10 mM NH_4OAc in 10% H_2O, 90% CH_3CN) in a linear gradient. Clean separation of the isomers was not possible, so the trace as reported was used for further experiments.

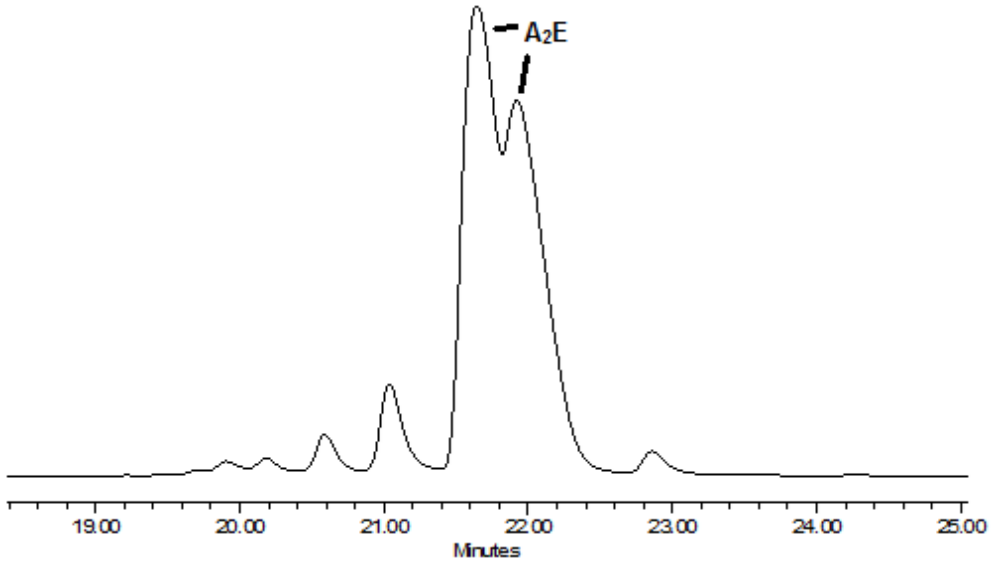


Figure 1.25 Reverse phase HPLC trace of the preparative DCC library for the synthesis of **A_2E**, monitored at 254nm.

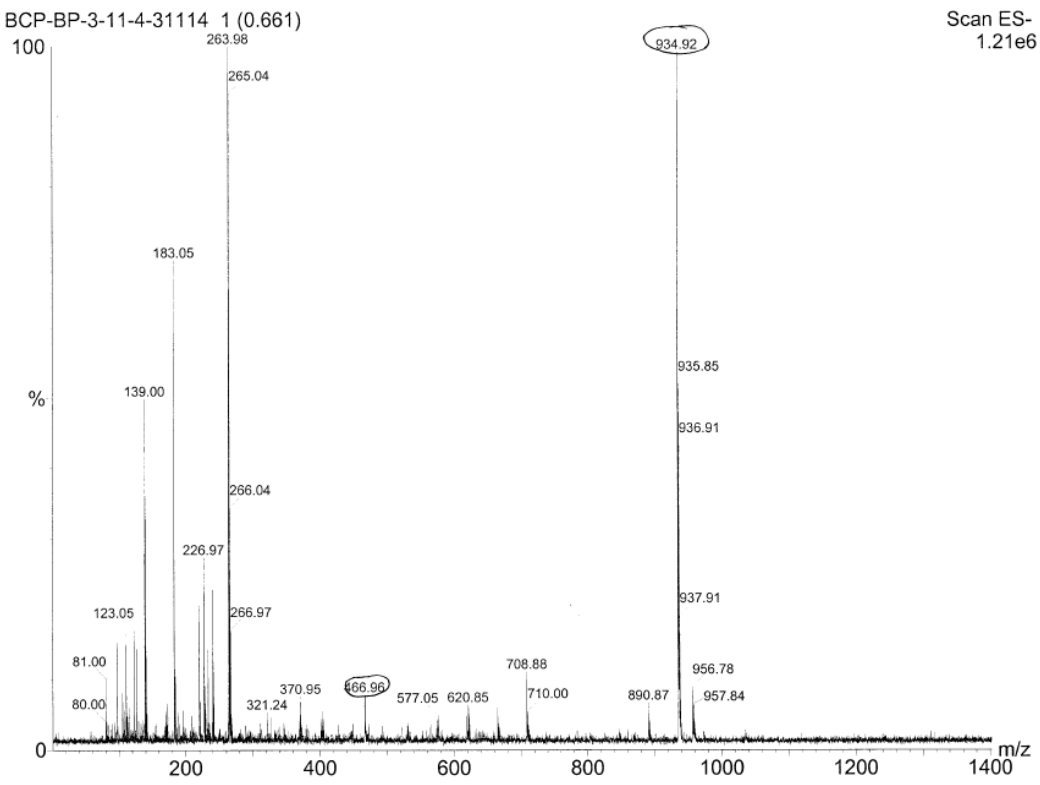


Figure 1.26 Mass spectrum of **A_2E** (-ESI). $[M-H]^{-1}$ at 934.92 and $[M-2H]^{-2}$ at 466.96.

1.4.4 Extinction Coefficient Determination

In order to have reproducible concentrations of receptor, extinction coefficients were determined using mixtures of isomers from **A_2E**. After purification using NH_4OAc buffered solvents, the receptors were lyophilized for at least one week to ensure removal of the volatile NH_4Oac salts. The dry sample was then taken up into anhydrous methanol and filtered with a 0.33 um filter into a tared vial. The methanol was evaporated and then the sample further dried under vacuum. After determining the mass, the dry sample was dissolved in 10 mM sodium borate buffer, pH 8.5 and diluted to 1 mM. Serial dilutions were performed to give ten samples which were analyzed at a variety of wavelengths. The extinction coefficient for **A_2E** was 4,812 M^{-1} cm^{-1} at 325 nm.

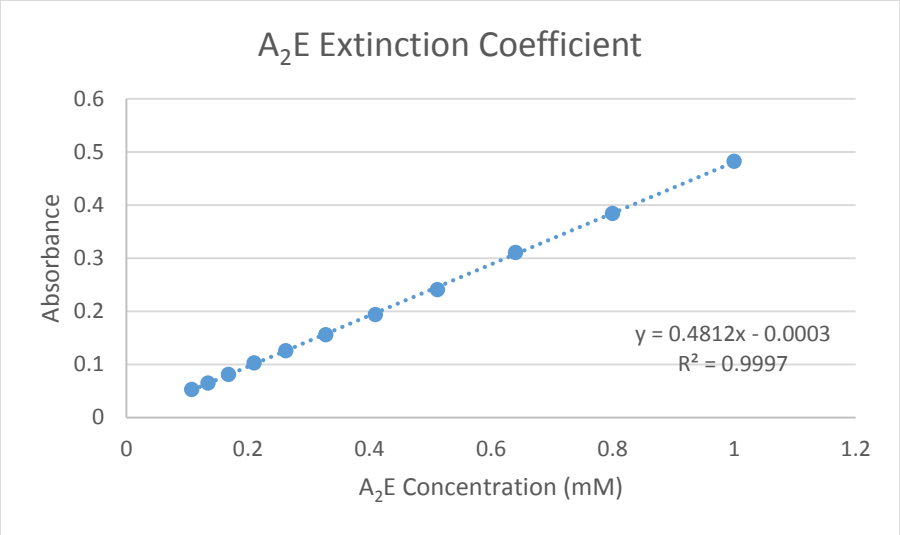


Figure 1.27 Extinction coefficient determination of **A_2E**. The extinction coefficient was determined as the slope of the linear regression

1.4.5 Isothermal Titration Calorimetry Binding Experiments

Depending on the system studied, ITC titrations were performed with a range of ~0.7-2 mM solution of peptide into ~60-200 uM of receptor. For weaker interactions, the c-value is low, so there is a higher degree of error. While one-site binding is assumed, n-values do deviate from 1, which can be attributed to both the error in accurately determining receptor concentration.

Heat of dilution titrations were measured on a Microcal AutoITC200 at 298K. In parallel to binding measurements by ITC, peptide (1-2 mM) was titrated into sodium borate buffer (10 mM, pH 8.5) using 2.0 μL increments every 3 minutes. The resulting data was manually integrated to reduce error in automatic baseline calculations. The resulting normalized changes in enthalpy (NDH) measurements were normalized for peptide concentration and subtracted directly from NDH measurements for all subsequent ITC titrations from that peptide stock solution.

Table 1.3 Thermodynamic binding data obtained for binding of the receptors to Ac-WGGG-QTARK(Me)nSTG-NH2 as measured by ITC.[a]

Entry	Receptor [b]	Peptide	K_d [c] (μM)	S. Factor [d]	ΔG [c] (kcal/mol)	ΔH [c] (kcal/mol)	TΔ (kcal
1 [e]	**A_2B**	KMe3	2.6 ± 0.1	-	-7.63 ± 0.03	-11.26 ± 0.05	-3.61
2 [e]	**A_2B**	KMe2	6.3 ± 0.3	2.4	-7.10 ± 0.07	-11.65 ± 0.09	-4.5
3 [e]	**A_2B**	KMe	13.9 ± 0.1	5.4	-6.64 ± 0.01	-9.65 ± 0.06	-3.00
4 [e]	**A_2B**	Lys	22 ± 1	8.3	-6.38 ± 0.02	-9.2 ± 0.2	-2.9
5 [e]	**A_2B**	R8GKMe3	17.1 ± 0.1		-6.52 ± 0.01	-12.37 ± 0.01	-5.84
6	**A_2C**	KMe3	2.3 ± 0.1	-	-7.69 ± 0.02	-10.4 ± 0.1	-2.7
7	**A_2C**	KMe2	2.8 ± 0.2	1.2	-7.57 ± 0.04	-9.52 ± 0.08	-1.9
8	**A_2C**	KMe	13.8 ± 0.7	6.0	-6.63 ± 0.03	-10.25 ± 0.01	-3.61
9	**A_2C**	Lys	22 ± 1	9.6	-6.34 ± 0.03	-10.1 ± 0.4	-3.7
10	**A_2C**	R8GKMe3	29 ± 3		-6.17 ± 0.05	-12.1 ± 0.4	-5.9
11 [f]	**A_2E**	KMe3	0.191 ± 0.002	-	-9.16 ± 0.01	-14.0 ± 0.3	-4.8
12 [f]	**A_2E**	KMe2	0.5 ± 0.1	2.6	-8.5 ± 0.1	-12.67 ± 0.05	-4.1
13 [f]	**A_2E**	KMe	1.6 ± 0.2	8.4	-7.92 ± 0.08	-11.92 ± 0.07	-3.9
14 [f]	**A_2E**	Lys	6.7 ± 0.1	35	-7.05 ± 0.01	-11.26 ± 0.02	-4.16
15 [f]	**A_2E**	R8GKMe3	2.7 ± 0.3		-7.59 ± 0.06	-12.2 ± 0.1	-4.5
16 [e]	**A_2N**	KMe3	0.30 ± 0.04	-	-8.91 ± 0.07	-12.0 ± 0.5	-3.1
17 [e]	**A_2N**	KMe2	4.1 ± 0.5	14	-7.36 ± 0.04	-12.5 ± 0.4	-5.1
18 [e]	**A_2N**	KMe	40 ± 4	130	-6.01 ± 0.06	-12.0 ± 0.5	-6.0
19 [e]	**A_2N**	Lys	10.5 ± 0.9	35	-6.80 ± 0.05	-7.3 ± 0.3	-0.5
20 [e]	**A_2N**	R8GKMe3	1.3 ± 0.2		-8.05 ± 0.08	-13.4 ± 0.5	-5.3
21 [f]	**A_2G**	KMe3	1.4 ± 0.1	-	-8.00 ± 0.05	-9.9 ± 0.2	-1.8
22 [f,g]	**A_2G**	KMe2	13.2 ± 2.4	10	-6.6 ± 0.1	-11.5 ± 0.1	-4.9
23 [f]	**A_2G**	KMe	15 ± 1	11	-6.57 ± 0.04	-10.2 ± 0.4	-3.6
24 [f,h]	**A_2G**	Lys	>58	>40	< -5.8	-	
25 [f,g]	**A_2G**	R8GKMe3	5.4 ± 0.1		-7.19 ± 0.01	-9.0 ± 0.1	-1.8

(a) All data determined by ITC, fit to one-site binding model; Conditions: 26 °C, in 10 mM sodium borate buffer, pH 8.5. (b) All receptors are mixtures of isomers except *rac*-**A_2B** and *meso*-**A_2N**. (c) Errors are from averages of three trials, unless noted otherwise. (d) S. factor is selectivity, which is calculated as the fold difference in affinity for KMe3 over the designated methylation state of the peptide in that row. (e) Data reported by Pinkin and Waters.[7] (f) Average of two trials. (g)Error determined by propagation from curve fitting and averages. (h)These values are approximate because the c-value for these experiments was <1.

1.4.5.1 A_2E Example Titrations

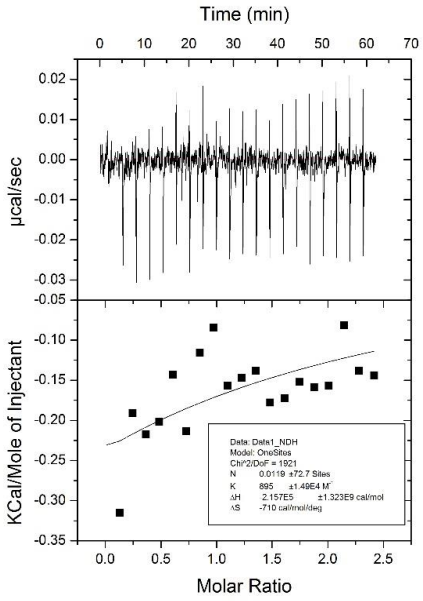


Figure 1.28. ITC of R8GK9G (Ac-WGGG-QTAGGSTG-NH_2) (1.2 mM) into **A_2E** (120 uM) at 26°C in 10 mM borate buffer, pH 8.5.

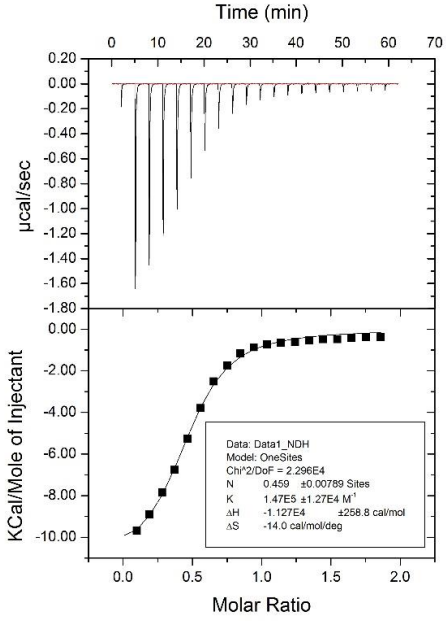


Figure 1.29 One of two trials of Lys (Ac-WGGG-QTARKSTG-NH_2) (0.996 mM) into **A_2E** (109 uM) at 26°C in 10 mM borate buffer, pH 8.5.

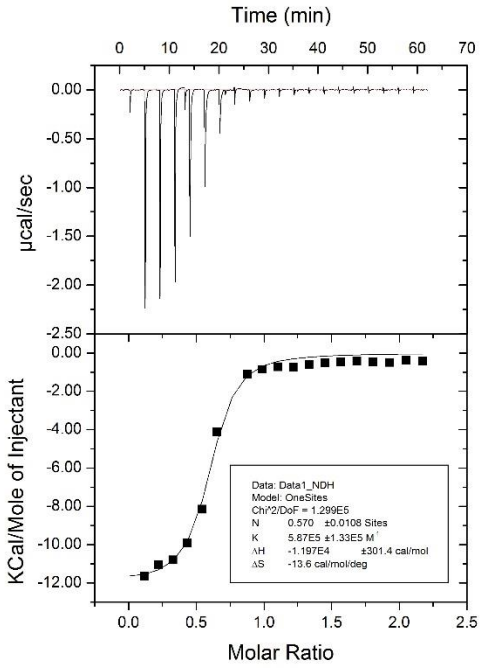


Figure 1.30 One of two trials of KMe (Ac-WGGG-QTARKMeSTG-NH_2) (1.1 mM) into **A_2E** (103 uM) at 26°C in 10 mM borate buffer, pH 8.5.

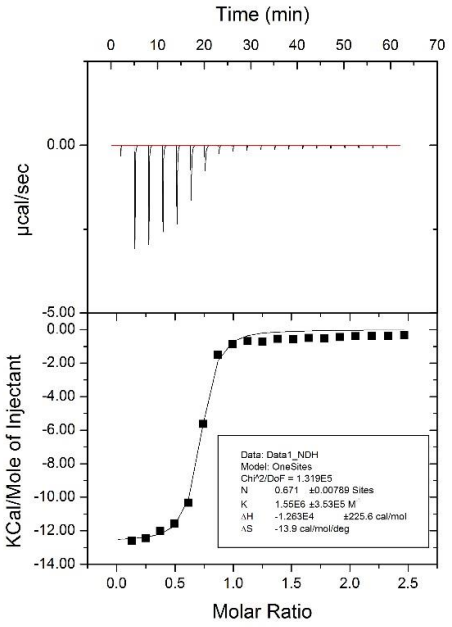


Figure 1.31 One of two trials of KMe2 (Ac-WGGG-QTARKMe$_2$STG-NH_2) (1.2 mM) into **A_2E** (99 uM) at 26°C in 10 mM borate buffer, pH 8.5.

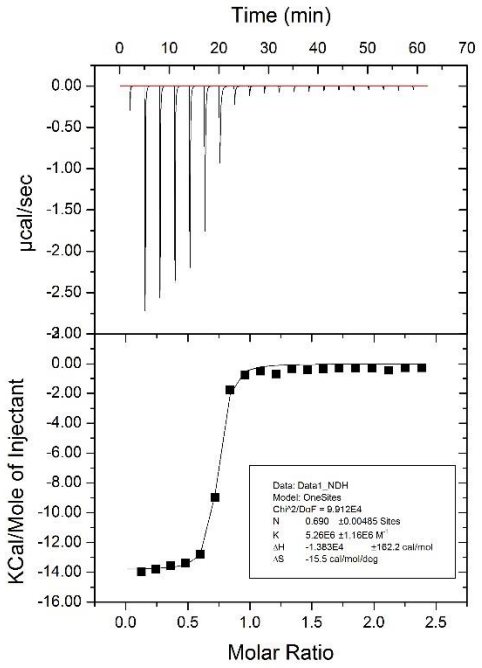


Figure 1.32 One of two trials of KMe3 (Ac-WGGG-QTARKMe$_3$STG-NH$_2$) (0.95 mM) into **A$_2$E** (80 uM) at 26°C in 10 mM borate buffer, pH 8.5.

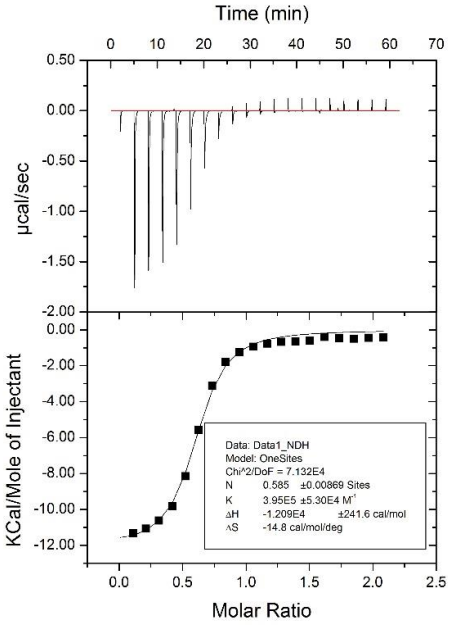


Figure 1.33 One of two trials of R8GKMe3 (Ac-WGGG-QTAGK(Me)$_3$STG-NH$_2$) (1.001 mM) into **A$_2$E** (98 uM) at 26°C in 10 mM borate buffer, pH 8.5.

REFERENCES

(1) Annuziato, A. *Nature Education.* **2008**, *1*, 26.

(2) McGinty, R. K.; Tan, S. *Chem. Rev.* **2015**, *115*, 2255.

(3) Luger, K.; Collins, F. *Life Sci.* **2001**, 1.

(4) Braunstein, M.; Sobel, R. E.; Allis, C. D.; Turner, B. M.; Broach, J. R. *Mol. Cell. Biol.* **1996**, *16*, 4349.

(5) Jenuwein, T.; Allis, C. D. *Science* **2001**, *293*, 1074.

(6) Kouzarides, T. *Cell* **2007**, *128*, 693.

(7) Falkenberg, K. J.; Johnstone, R. W. *Nat. Rev. Drug Discov.* **2014**, *2*, 673.

(8) Shiio, Y.; Eisenman, R. N. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13225.

(9) Cao, J.; Yan, Q. *Front. Oncol.* **2012**, *2*, 1.

(10) Bowman, G. D.; Poirier, M. G. *Chem. Rev.* **2015**, *115*, 2274.

(11) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838.

(12) He, Y.; Korboukh, I.; Jin, J.; Huang, J. *Acta Biochim. Biophys. Sin.* **2012**, *44*, 70.

(13) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025.

(14) Yun, M.; Wu, J.; Workman, J. L.; Li, B. *Cell Res.* **2011**, *21*, 564.

(15) Eisert, R. J.; Waters, M. L. *ChemBioChem* **2011**, *12*, 2786.

(16) Guo, Y.; Nady, N.; Qi, C.; Allali-Hassani, A.; Zhu, H.; Pan, P.; Adams-Cioaba, M. a.; Amaya, M. F.; Dong, A.; Vedadi, M.; Schapira, M.; Read, R. J.; Arrowsmith, C. H.; Min, J. *Nucleic Acids Res.* **2009**, *37*, 2204.

(17) Trojer, P.; Li, G.; Sims, R. J.; Vaquero, A.; Kalakonda, N.; Boccuni, P.; Lee, D.; Erdjument-Bromage, H.; Tempst, P.; Nimer, S. D.; Wang, Y. H.; Reinberg, D. *Cell* **2007**, *129*, 915.

(18) Gao, C.; Herold, J. M.; Kireev, D.; Wigle, T.; Norris, J. L.; Frye, S. *J. Am. Chem. Soc.* **2011**, *133*, 5357.

(19) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books: California, 2006.

(20) Southall, N. T.; Dill, K. a.; Haymet, a. D. J. *J. Phys. Chem. B* **2002**, *106*, 521.

(21) Meyer, E. E.; Rosenberg, K. J.; Israelachvili, J. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15739.

(22) Meyer, E. a.; Castellano, R. K.; Diederich, F. *Angew. Chem. Int. Ed.* **2003**, *42*, *1210*.

(23) Biedermann, F.; Nau, W. M.; Schneider, H.-J. *Angew. Chem. Int. Ed.* **2014**, 2.

(24) Biedermann, F.; Uzunova, V. D.; Scherman, O. a.; Nau, W. M.; De Simone, A. *J. Am. Chem. Soc.* **2012**, *134*, 15318.

(25) Ma, J. C.; Ma, J. C.; Dougherty, D. A.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303.

(26) Sunner, J.; Nishizawa, K.; Kebarle, P. *J. Phys. Chem.* **1981**, *85*, 1814.

(27) Meot-Ner, M.; Deakyne, C. a. *J. Am. Chem. Soc.* **1985**, *107*, 474.

(28) Meot-Ner, M.; Deakyne, C. a. *J. Am. Chem. Soc.* **1985**, *63*, 469.

(29) Gallivan, J. P.; Dougherty, D. a. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459.

(30) Gallivan, J. P.; Dougherty, D. a. *J. Am. Chem. Soc.* **2000**, *122*, 870.

(31) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2009**, *5*, 2301.

(32) Stauffer, D. a.; Dougherty, D. a. *Tetrahedron Lett.* **1988**, *29*, 6039.

(33) Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11184.

(34) Gamal-Eldin, M. a; Macartney, D. H. *Org. Biomol. Chem.* **2013**, *11*, 488.

(35) Assaf, K. I.; Nau, W. M. *Chem. Soc. Rev.* **2015**, *44*, 394.

(36) Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. *ChemBioChem* **2010**, *11*, 63.

(37) Daze, K. D.; Ma, M. C. F.; Pineux, F.; Hof, F. *Org. Lett.* **2012**, *14*, 1512.

(38) Daze, K. D.; Pinter, T.; Beshara, C. S.; Ibraheem, A.; Minaker, S. a.; Ma, M. C. F.; Courtemanche, R. J. M.; Campbell, R. E.; Hof, F. *Chem. Sci.* **2012**, *3*, 2695.

(39) Tabet, S.; Douglas, S. F.; Daze, K. D.; Garnett, G. a E.; Allen, K. J. H.; Abrioux, E. M. M.; Quon, T. T. H.; Wulff, J. E.; Hof, F. *Bioorganic Med. Chem.* **2013**, *21*, 7004.

(40) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discov. Today* **2002**, *7*, 117.

(41) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 321.

(42) Lehn, J.-M. *Chem. A Eur. J.* **1999**, *5*, 2455.

(43) Wessjohann, L. a.; Rivera, D. G.; León, F. *Org. Lett.* **2007**, *9*, 4733.

(44) Van Gerven, P. C. M.; Elemans, J. a a W.; Gerritsen, J. W.; Speller, S.; Nolte, R. J. M.; Rowan, A. E. *Chem. Commun.* **2005**, *28*, 3535.

(45) Gagné, M. R.; Ghosh, S.; Ingerman, L. a.; Frye, A. G.; Lee, S. J.; Waters, M. L. *Org. Lett.* **2010**, *12*, 1860.

(46) Black, S. P.; Sanders, J. K. M.; Stefankiewicz, A. R. *Chem. Soc. Rev.* **2014**, *43*, 1861.

(47) Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Chem. - A Eur. J.* **2008**, *14*, 2153.

(48) Ingerman, L. a; Cuellar, M. E.; Waters, M. L. *Chem. Commun.* **2010**, *46*, 1839.

(49) Pinkin, N. K.; Waters, M. L. *Org. Biomol. Chem.* **2014**, *12*, 7059.

(50) Vial, L.; Ludlow, R. F.; Leclaire, J.; Pérez-Fernández, R.; Otto, S. *J. Am. Chem. Soc.* **2006**, *128*, 10253.

(51) Beaver, J.; Peacor, B.; Bain, J.; James, L.; Waters, M. *Org. Biomol. Chem.* **2015**, *13*, 3220.

(52) Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. a. *J. Am. Chem. Soc.* **1999**, *121*, 1192.

CHAPTER 2 Fluorescence Assay for Enzymatic Methylation of Histone Peptides

2.1 Introduction

The methylation of lysine and arginine plays a critical role in regulating the genetic landscape, but also presents unique challenges for study. Many methods exist for detecting and characterizing methylation, but a convenient, rapid, and universal label-free approach is still lacking. Recent work has developed supramolecular indicator displacement based approaches for the sensing of enzymatic reactions, including trimethylation of lysine, and we believe that this approach is easily adaptable to the study of other methylation events. Through a competitive fluorescence displacement assay, we can leverage our knowledge of supramolecular analyte binding to create a rapid, versatile, and label-free assay to study enzymatic reactions of high importance.

Previously, we have described the application of DCC towards the synthesis and characterization of novel receptors for the methylated forms of lysine.[1] Combined with the receptors discovered by previous members of the lab, we have established a suite of small-molecule hosts for methylated lysine and arginine, each of which displays unique binding affinities for the various analytes.[2–4] Using these receptors, we were able to screen a variety of previously reported environmentally sensitive fluorophores to create a turn-on fluorescence assay for both lysine and arginine methylation. Upon characterization of the receptor/dye pair, we attempted to optimize the reaction conditions for the lysine methyltransferase enzyme G9a, balancing the requirements for both assay sensitivity as well as enzyme activity, and were able to

transfer this system to the initial characterization of the protein arginine methyltransferase (PRMT) family of enzymes.

2.2 Background

2.2.1 Protein Lysine Methyltransferases (PKMTs)

The post-translational methylation of lysine encodes a diverse number of downstream biological events. Methylation is highly context dependent, with different sites conferring either transcriptional activation or repression. Additionally, the degree of methylation can have different outcomes, even conserved within the same position in the histone tail.[5] For example, the methylation of lysine 9 on the histone H3 tail is responsible for the X-chromosome inactivation when dimethylated or increased transcriptional activation when monomethylated.[6,7] Because of its importance, lysine methylation is highly regulated, with separate enzymes capable of both installing the varying degrees of methylation, known as 'writers', as well as other enzymes responsible for removing the modification, or 'erasers'.[8] Misregulation has been implicated in a large number of cancers.[9–12] As more is understood about the methylation event in a biological context, there is heavy interest in understanding the 'writers' responsible for its installation.

The methylation of lysine is catalyzed by a large family of enzymes called protein lysine methyltransferases, or PKMTs, shown in Figure 2.1.

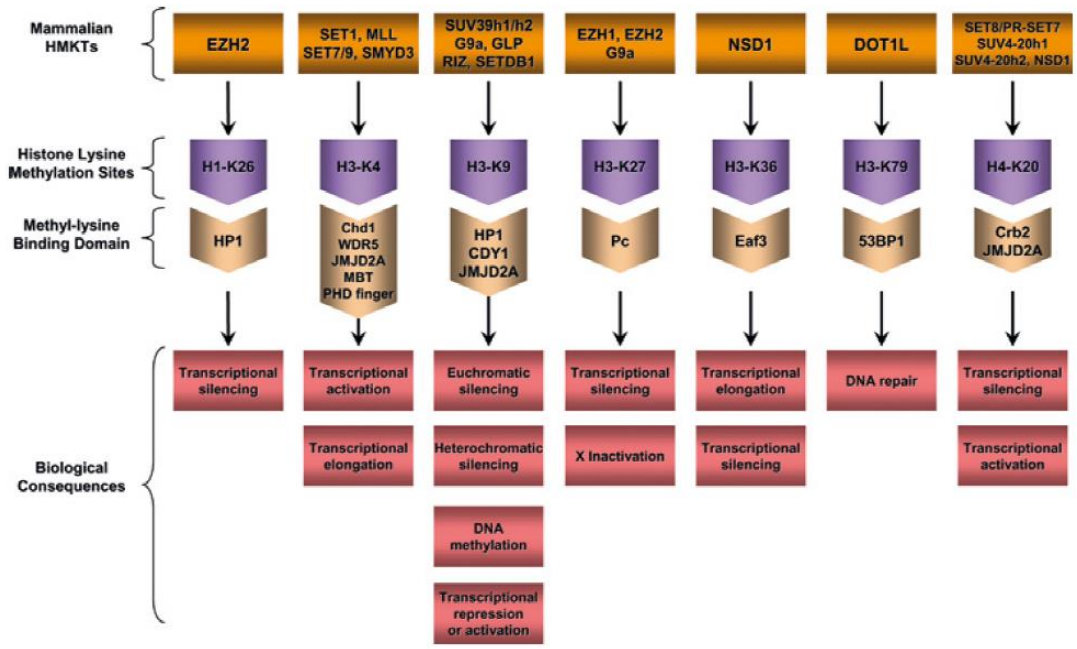


Figure 2.1 Family of PKMTs and their respective histone targets. Each methylation is responsible for a unique biological function dependent on the site and degree of methylation.[13] Reprinted with permission from *Cell Mol. Life Sci.*, 2006, *63*, 2755-2763.

Each member of this family, save one, contains a conserved SET-domain responsible for catalyzing the methylation.[13] The single exception is the enzyme Dot1, which catalyzes the methylation of H3-K79, a lysine residue in the histone fold, which requires a unique method of catalysis due to being more challenging to access than the free N-terminal tails.[14] Regardless, Dot1 and the SET family all catalyze a similar reaction, the transfer of a methyl group from the sulfur of S-Adenosylmethionine to the ε-amino group of lysine, as shown in Figure 2.2.

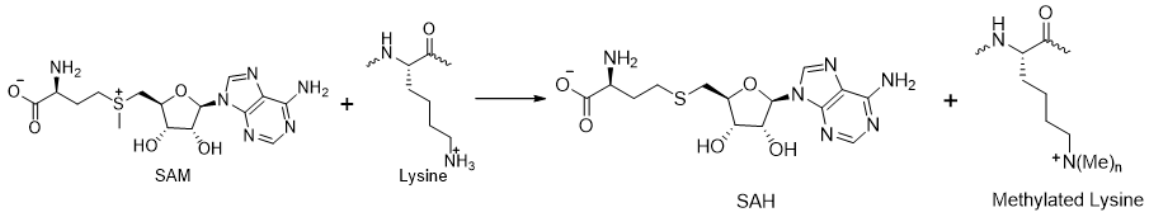


Figure 2.2 General reaction scheme for the methylation of lysine by PKMT enzymes and the cofactor S-Adenosylmethionine.

Inside the active site of the SET-domain are two key binding sites, one each for the lysine substrate and the SAM cofactor. Connecting these sites is a hydrophobic channel responsible for positioning the lysine residue for nucleophilic attack onto the methyl group of SAM.[13] This channel is also responsible for the processive nature of the enzymes, allowing the installation of the required degree of methylation prior to protein dissociation.[15] A crystal structure of the PKMT G9a bound to both its dimethylated H3K9 product and the SAH cofactor within this hydrophobic channel is shown in Figure 2.3.

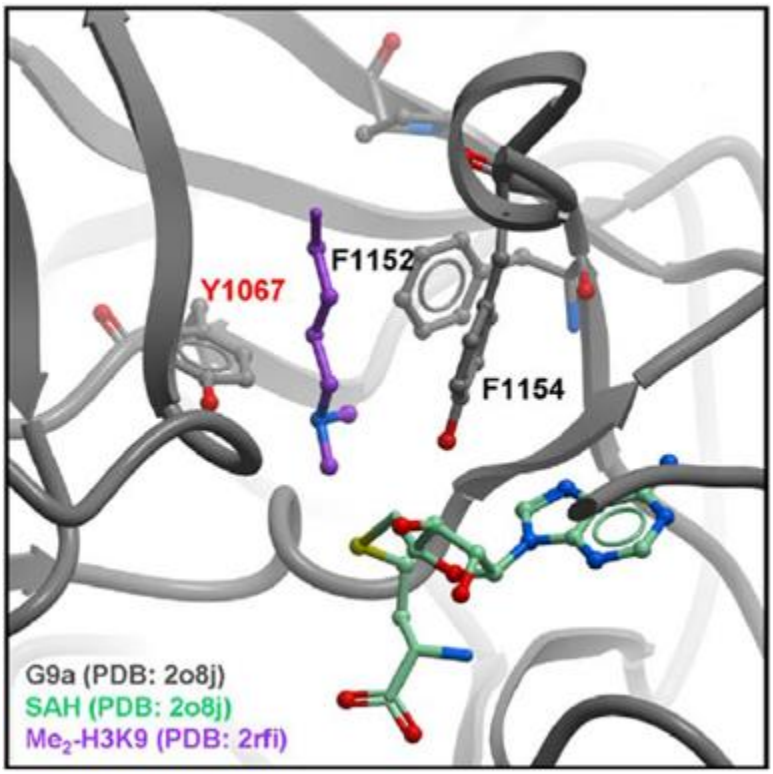


Figure 2.3 Model of G9a (grey) bound to H3K9$(Me)_2$ (purple) and SAM cofactor (green). Key residues of the enzyme active site are shown.[16]

The H3K9 residue binds to G9a and is positioned in the channel by two tyrosine residues and a phenylalanine. Tyrosine 1067 is proposed to hydrogen bond to the ε-amino group, disfavoring it from obtaining the correct orientation to allow for a third methylation event.[16]

SAM has been shown to bind in a compact, bent form, positioning the methyl group opposite of lysine in the channel.[17] This bent conformation is reinforced by a large number of non-covalent interactions within the binding site and is critical to enzyme activity. Additionally, the enzyme is known to have an acidic groove responsible for making contacts with the neighboring sequence of the histone H3 tail and guiding the enzyme to its substrate. This groove is also thought to play a role in the biological regulation of the methyltransferase, with neighboring phosphorylation significantly lowering enzyme activity, most likely due to a disruption of the favorable peptide binding.[16,18]

Recent studies have been conducted to determine the mode of selectivity of the PKMTs. Because the degree of methylation is critical to the correct function and the enzymes are processive, there must be a mechanism for stopping the methylation at the correct point. The first step of the enzymatic catalysis post substrate binding is the deprotonation of lysine to make the active species. There have been several proposed mechanisms for this reaction, the first of which is a conserved tyrosine residue inside the SET-domain. In the SET7/9 dimethyltransferase enzyme, this tyrosine was suggested to be acting as a general base over the course of the processive reaction.[19] The tyrosine was first deprotonated by a bulk water before proton transfer from the ε-amino group of lysine could occur. In 2008, the Bruice group proposed a different mechanism for the bulk water channel, shown in Figure 2.4.[20]

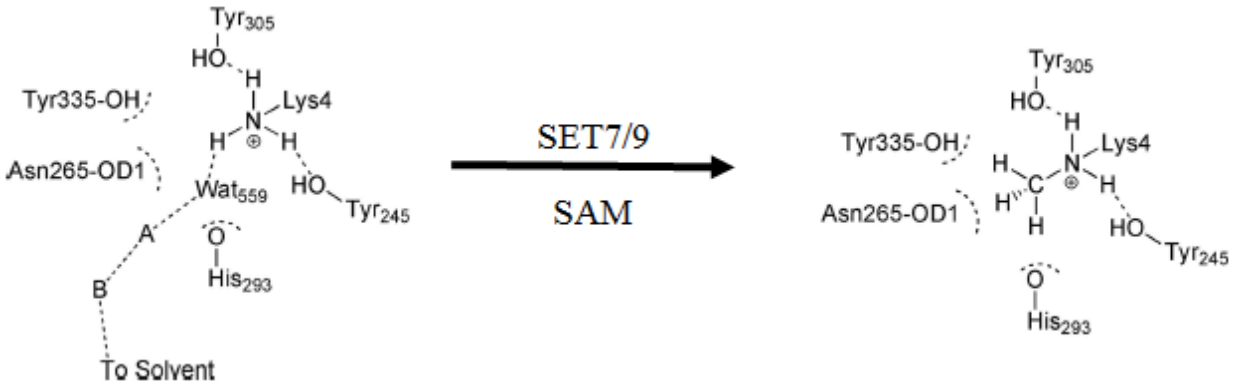


Figure 2.4 MD Simulations of the water channel promoted deprotonation of H3K4 and subsequent methylation in the SET7/9 active site.[20] Copyright 2008 PNAS.

In the simulation, they observed that there was a route for bulk water to reach the ε-amino of lysine, and the resulting deprotonation allowed the methylation reaction to occur. Upon methylation, the water channel became blocked, preventing further deprotonation and resulting in selective monomethylation. The study also did simulations on PKMT vSET which catalyzes the trimethylation of H3K27. There they observed two separate water channels that could perform three deprotonation events prior to protein dissociation, furnishing the trimethylated product.[20]

2.2.2 Protein Arginine Methyltransferases (PRMTs).

The methylation of arginine, similar to the methylation of lysine, is catalyzed by the PRMT family of enzymes and is carried out by the transfer of a methyl group from SAM to a guanidino nitrogen on the arginine side chain.[21] There are three methylated states of arginine, but unlike lysine the higher order methylations are not processive. Rather, there are two main classes of PRMTs capable of higher degrees of methylation, the type I PRMT which catalyze formation of asymmetric dimethyl arginine, $R_a(Me)_2$, or the type II PRMT which forms symmetric dimethyl arginine, $R_s(Me)_2$, as shown in Figure 2.5.

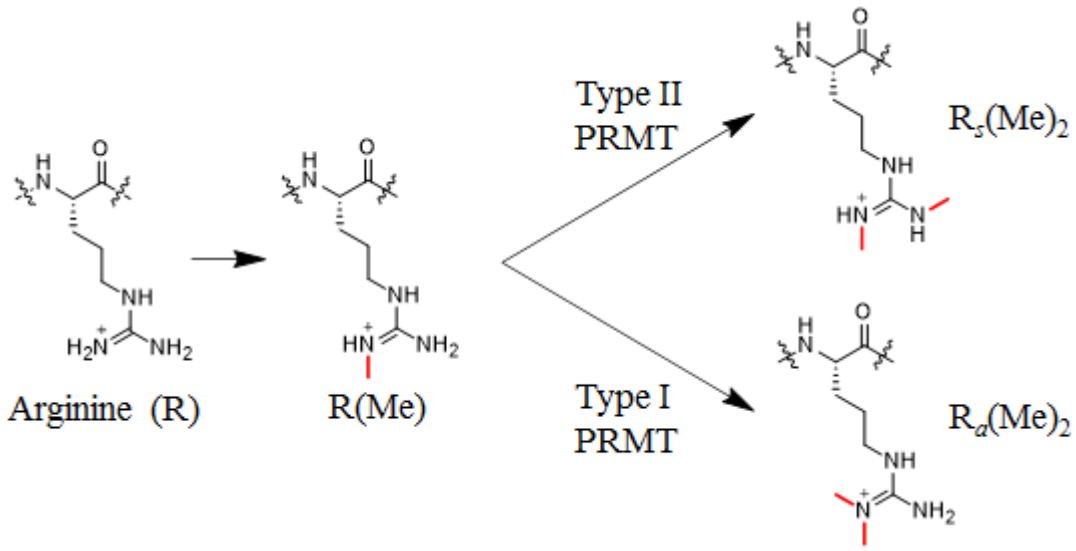


Figure 2.5 Methylation of arginine catalyzed by two distinct families of enzyme. Asymmetric dimethylation is catalyzed by the Type I PRMT whereas symmetric dimethylation is carried out by the Type II PRMT.

Similar to lysine methylation, arginine methylation is responsible for a wide number of biological functions.[22] In the histone H4 tail, arginine 3 can be dimethylated both asymmetrically and symmetrically by PRMT1 and PRMT5 respectively, but each methylation state is responsible for a different event, with PRMT1 activating transcription and PRMT5 silencing it. Additionally, arginine methylation can act in tandem with other histone PTMs. In the context of $H3R2_a(Me)_2$, the Yang group reported that the recombination activating gene (RAG) 2 has a PHD finger which binds to $H3K4(Me)_3$, but that this binding is enhanced by the neighboring methylated arginine.[23] This suggests that arginine methylation, like lysine methylation, is part of a complex landscape of regulatory modifications.

All PRMT enzymes share a catalytic core comprising a Rossmann fold, responsible for the SAM binding domain, and an acidic β-barrel for substrate binding due to the multiple basic residues in typical PRMT substrates, shown in Figure 2.6.[24,25]

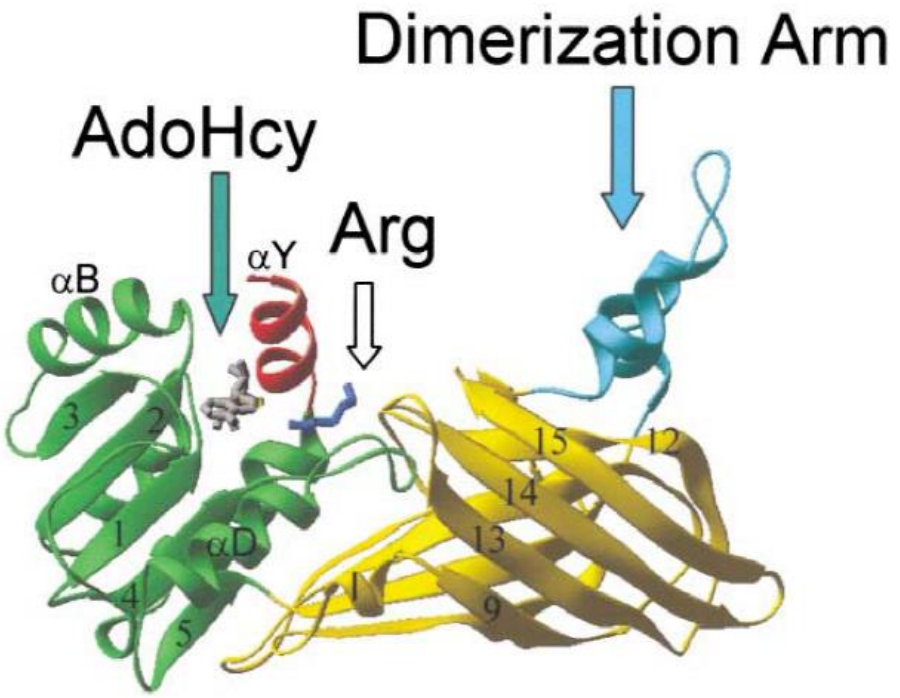


Figure 2.6 Monomeric structure of PRMT1. SAM binding domain is shown in green, with SAH (AdoHcy) shown in grey. The β-barrel is shown in yellow with bound arginine shown as the dark blue stick structure. The dimerization region of PRMT is highlighted in light blue.[25] Reprinted from *Structure, 11*, Zhang, X., Cheng, X., Structure of the Predominant Protein Arginine Methyltransferase PRMT1 and Analysis of Its Binding to Substrate Peptides. 509-520. Copyright 2003, with permission from Elsevier.

Interestingly, there is also a hydrophobic region responsible for dimerization that is conserved across the PRMT family, and this dimerization is critical for enzymatic activity, with mutation of the region's residues to alanine completely abolishing activity in the yeast arginine methyltransferase.[26] Recent studies have also shown that while the type I methyltransferases such as PRMT1 are known to be at least partially processive, the type II PRMT5 is distributive, releasing monomethyl arginine completely before re-binding and catalyzing the second, symmetric methylation.[27,28] Unlike the lysine methyltransferases, PRMT1 is proposed to catalyze the nucleophilic transfer of the S-methyl group from SAM by going first through a dication intermediate followed by proton transfer to an interior glutamic acid.[27]

Both lysine methylation and arginine methylation are critically important modifications for gene regulation. There has been a large focus in recent years on the inhibition of methyltransferases as potential drug candidates.[29] In that time, several small molecule inhibitors have been discovered, including UNC0638 a nanomolar inhibitor selective for the lysine methyltransferases G9a and GLP, as well as SGC707, a selective PRMT3 inhibitor.[30,31] As more study is devoted to these enzymes, there remains a need for effective tools to determine their mechanisms and activities. This becomes especially true as more is learned about the vast complexity of the 'histone code', the language of modifications in which enzymes recognize not only their specific substrate but the landscape of neighboring modifications as well.[32,33] These complex substrates and large variety of enzyme pathways are prompting the development of new rapid and high-throughput assays for characterizing methylation.

2.2.3 Methylation Assays

2.2.3.1 Current Tools

Currently there are several methods to monitor methyltransferase enzymes. These include radioactive assays, antibody based approaches, MS/electrophoresis, and reaction product coupled fluorescence assays.[34] Radioactive assays monitor the incorporation of a radioactive methyl group from SAM, normally in the form of tritium labeling. Because of this, the assays are highly sensitive, though expensive with byproducts that are hard to handle and dispose of.[35] Additionally, radioactive assays only give end-point data and cannot discriminate the degree of methylation. Antibody based approaches can solve these problems and have the advantage of being able to determine degree of methylation. Antibodies enable the facile study of both modifications and the genes they encode through a variety of chromatin immunoprecipitation (ChIP) methods. ChIP-seq utilizes the cross linking of DNA to its neighboring histone proteins

through formaldehyde incubation followed by sample digestion. The resulting fragments can then be pulled out of solution with high-affinity antibodies for the methylated residue of interest and then the resulting genes can be sequenced.[36] This enables the isolation of only those histone-gene complexes that contain the target modification. Additionally, antibodies have been developed for a variety of high-throughput assays, including AlphaLisa, in which dye conjugate antibodies are used to quantify histone methylation.[37] However, antibodies for specific modifications have been found to be promiscuous and influenced by neighboring modifications, in addition to suffering from significant batch to batch reproducibility issues.[38,39] Because of this, methods are currently being developed to decouple the detection step of methyltransferase assays from the histone substrates.

Recently, assay development has turned to reaction byproduct coupled assays, in which the products generated by the enzyme, SAH in the case of the methyltransferases, can be rapidly and quantitatively detected to follow the reaction. In 2005, the Trievel group developed the SAH coupled assay described in Figure 2.7.

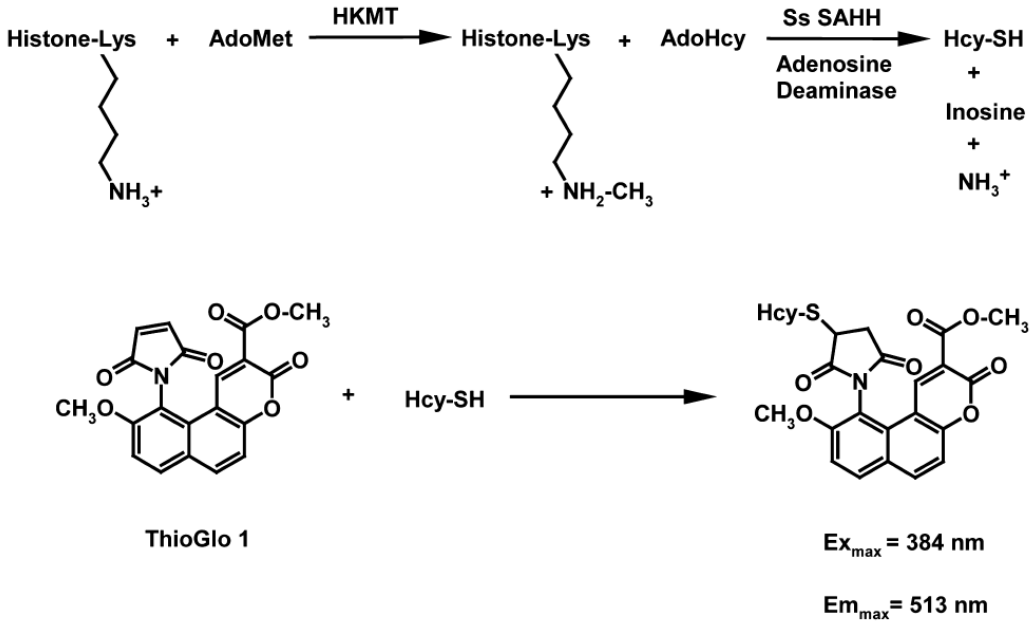


Figure 2.7 Reaction scheme for the SAH (AdoHcy) coupled lysine methyltransferase assay. Enzymatic assay is run in the presence of SAH hydrolase enzyme (Ss SAHH) in the presence of an adenosine deaminase. The resulting homocysteine (Hcy-SH) can be coupled to the fluorophore ThioGlo and quantified.[40] Reprinted from *Anal. Biochem.*, *342*, Collaza, E., Couture, J., Bulfer, S., Trievel, R., A Coupled Fluorescent Assay for Histone Methyltransferases. 86-92, Copyright 2005, with permission from Elsevier.

This assay is able to accurately determine kinetic parameters of a lysine methyltransferase enzyme quantitatively. The reaction is run in the presence of a SAH hydrolase, which forms adenosine and the reactive homocysteine. Adenosine is further deaminated to prevent the subsequent backwards recombination into SAH. The fluorescent assay allows aliquots to be removed and quenched followed by incubation with a fluorophore to furnish a kinetic assay for methylation.[40] However, this method of product coupled assays suffers from several drawbacks, including the optimization and handling of subsequent enzymatic reactions as well as potential for interference and required deconvolution of hits, which can complicate high-throughput analysis.

2.2.3.2 Indicator Displacement Assays

In light of the drawbacks of coupled assays, developments in fluorescence assays that rely on supramolecular complexation have shown promise for monitoring enzymatic reactions. These assays, termed supramolecular tandem assays, or indicator displacement assays (IDAs), utilize the competitive binding of both an analyte and an environmentally sensitive fluorophore, as shown in Figure 2.8.

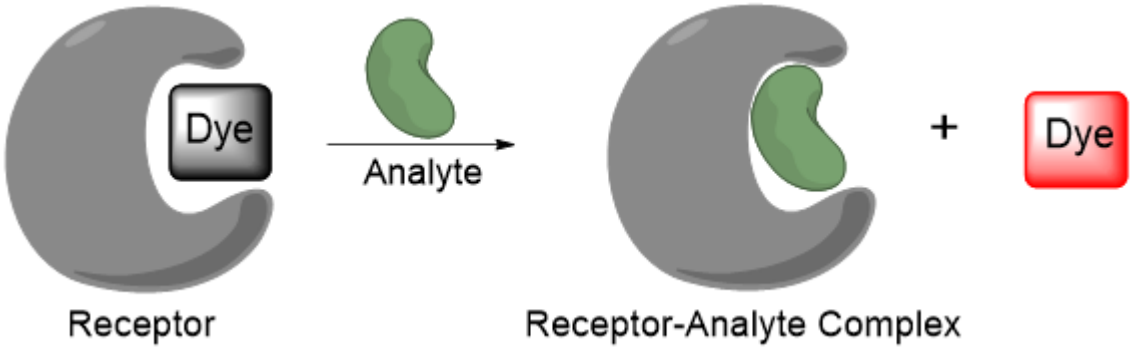


Figure 2.8 General scheme of a 'turn-on' indicator displacement assay, in which the fluorophore bound to receptor exists in a quenched state. Upon competitive binding, the dye exits the quenching environment, generating fluorescence in a concentration dependent manner.

The dyes chosen for these assays can be modulated by several mechanisms, including photoinduced electron transfer (PET), fluorescence resonance energy transfer (FRET), or even changes in local environment such as pH or polarity.[41] Additionally, the assay can operate in either a turn-on fashion, where analyte binding causes the dye to fluoresce, or turn-off, where binding quenches assay signal. The critical component of a functional IDA is the principle that host/dye binding affinity must be similar to host/analyte affinity.

The IDA gives several major advantages to a sensing application. Because it relies on supramolecular interactions, there is no need to tether the read-out to the sensor, allowing multiple indicators to be screened and even used with each receptor. Because the assay relies on analyte binding, studies have shown that the sensor can be both substrate sensitive as well as

product sensitive, providing additional operational flexibility for studying enzymatic reactions.[42] Additionally, while conditions need to be optimized based on binding parameters, there are no additional reaction conditions or steps that need to be developed or troubleshot, allowing a facile signal readout. It does have some drawbacks, including limitations in enzyme screening and inhibition assays if the cofactors or inhibitors are capable of binding to the receptor. However, because this competition can be easily controlled for, the IDA provides an easily adaptable platform for a variety of enzymatic reactions.

2.2.3.3 Enzyme Indicator Displacement Assays

IDAs have been used successfully in several different enzymatic activity assays. In 2009, the Nau group developed two IDA sensors to study the enzymatic reactions of arginase and diamine oxidase.[43] The reaction of diamine oxidase is an example of a substrate sensitive assay, in which the product is more weakly complexed to the host system, allowing the dye to enter and produce a fluorescence signal, as shown in Figure 2.9.

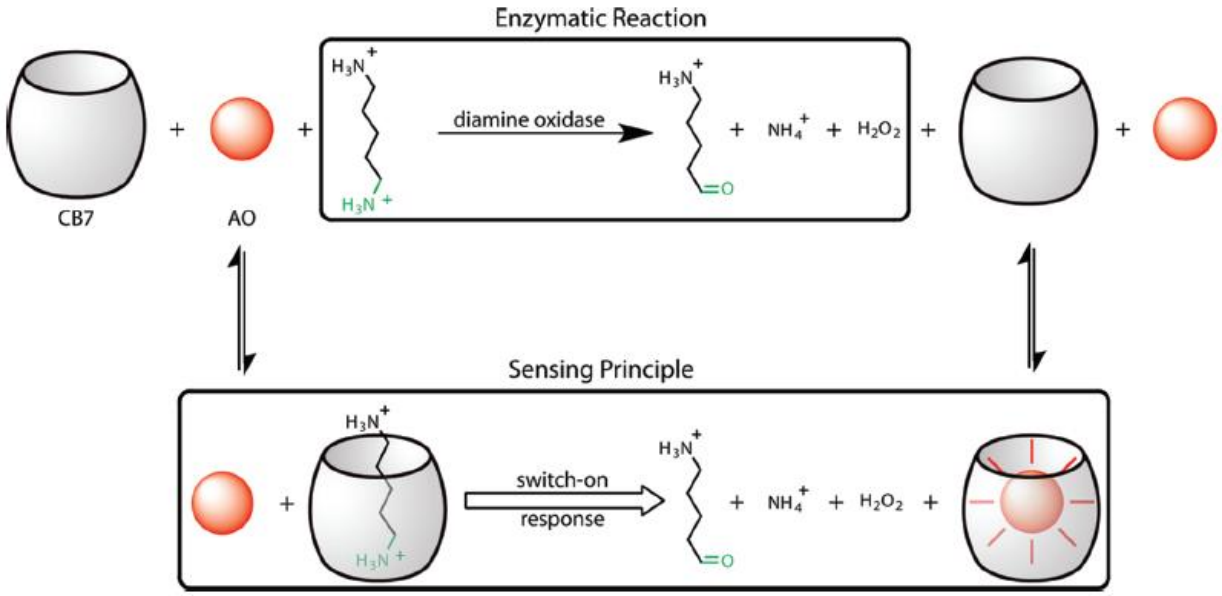


Figure 2.9 Diagram for the fluorescent diamine oxidase enzyme based on the CB7/AO sensor.[43] Reprinted with permission from *J. Am. Chem. Soc.*, 2009, *131*, 11558-11570. Copyright 2009 American Chemical Society.

In this system, the diamine cadaverine is favorably bound to CB7, a cucurbituril, due to a strong ion-dipole/hydrogen bonding interactions between the cationic ammonium groups and the electronegative oxygen rich rims. Upon enzymatic conversion of one amino group to the corresponding aldehyde, the binding affinity decreases by an order of magnitude. This allows the fluorophore acridine orange to bind to CB7, again through a positive interaction between cationic ammoniums, resulting in a fluorescence signal increase and determination of relative enzyme rates for the reaction.

In 2012, the Nau group expanded their displacement assays towards the study of the methyltransferase Dim-5, responsible for the trimethylation of lysine 9 on the H3 tail. Their sensor system utilized the host *p*-sulfonatocalix[4]arene (CX4) and the fluorophore lucigenin (LCG). Several groups have explored CX4 binding to methylated ammoniums, and have shown that CX4 is able to bind to trimethylated lysine without a strong interaction with the

unmethylated substrate, in the context of amino acids.[44] However, the addition of neighboring sequence to the lysine 9 increases the binding affinity for both the methylated and non-methylated species, reducing the selectivity. Because of this, the IDA exhibited only 30% fluorescence recovery from the enzymatic substrate. This small change was due in large part to the lower selectivity observed for $K(Me)_3$ over lysine, because the affinities of the longer peptides were similar. A lower receptor selectivity leads to lower difference in fluorophore displacement, lessening the dynamic range of the assay. However there was still enough signal in the assay to allow for the monitoring of the methylation reaction, which they could follow and analyze for apparent enzyme kinetics, as well as utilize to screen Dim-5 inhibitors.

2.2.4 Project Motivation

With the goal of establishing a sensitive label free assay for lysine and arginine methylation, we turned to development of an IDA using receptors developed in the Waters group. Previous work in the Waters lab identified receptors that are selective for trimethyl lysine, but still bind quite well to dimethyl lysine or asymmetric dimethyl arginine, both very important biological marks lacking high-throughput label-free assays, and shown in Figure 2.10, with binding affinities shown in Table 2.1.[2–4]

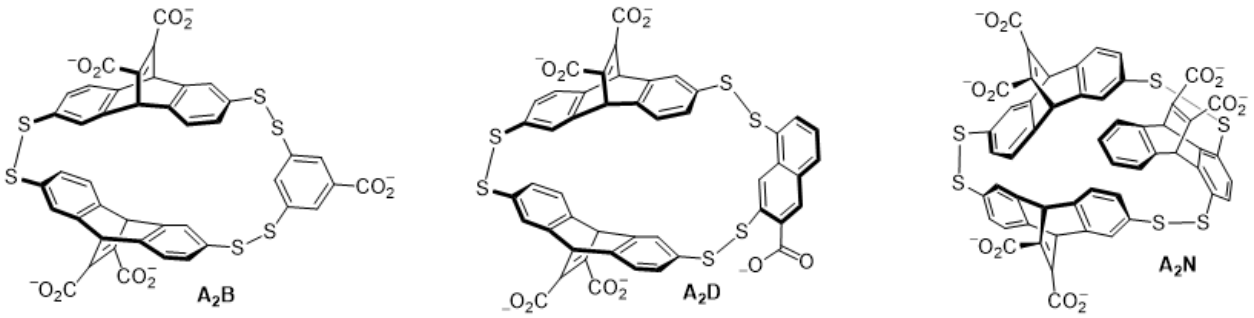


Figure 2.10 DCC receptors A_2B, A_2D, and A_2N.

Table 2.1 Binding affinity of A_2B, A_2D, and A_2N for methylated lysine measured by ITC.

Peptide	A_2B K_d (μM)[a]	A_2D K_d (μM)[b]	A_2N K_d (μM)[a]
Lys	21.9 ± 0.9	≥ 60	10.5 ± 0.9
Lys(Me)	13.9 ± 0.1	-	40 ± 4
$Lys(Me)_2$	6.4 ± 0.3	-	4.1 ± 0.5
$Lys(Me)_3$	2.6 ± 0.1	3.9 ± 0.5	0.3 ± 0.04

[a] Ac-WGGG-QTARK$(Me)_n$STG-NH_2 [b] Ac-YGG-QTARK$(Me)_n$STG-NH2

We hypothesized that, similar to CX4 and CB7, the receptors would exhibit strong binding affinities for a variety of organic fluorophores bearing quaternary ammoniums, and the hydrophobic aromatic pocket of the receptors could play a role in modulating the signal of these fluorophores, as shown in Figure 2.11.

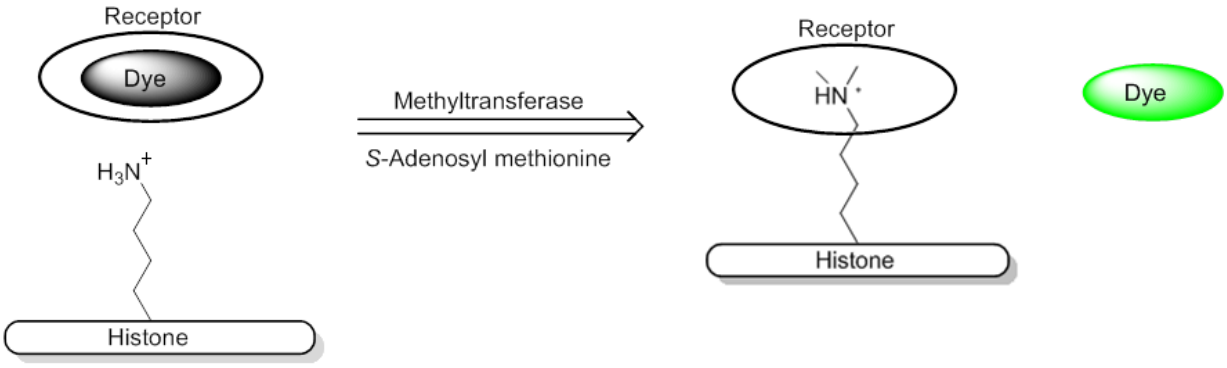


Figure 2.11 Diagram for the fluorescent sensing of the dimethyltransferase enzyme G9a. The proposed IDA is product sensitive due to increased binding affinity to the dimethylation mark of lysine.

The lysine methyltransferase G9a is responsible for dimethylating lysine 9 on the histone H3 tail, and its inhibition has been shown to limit the growth of several cancer cell lines.[30] Based on this scheme, and with the goal of studying the G9a, four key components required discovery and optimization. The first was selecting an environmentally sensitive fluorophore capable of

both binding to and being quenched by our receptors. Second, we required a receptor that when bound to the dye could successfully distinguish between a substrate unmethylated peptide and the dimethylated product. Finally, we had to optimize in tandem two different sets of conditions, those responsible for the best enzymatic activity and those for the best assay response. In doing so, we created a sensor assay that could be easily adaptable to a wide range of methyltransferases, beginning with lysine dimethylation and arginine asymmetric dimethylation.

2.3 Lysine Methyltransferase Assay

2.3.1 Fluorophore Screening

Our initial exploration of the indicator displacement assay was focused around the receptor **A_2B**, which was easily synthesized and purified in large quantities from preparative DCC libraries.[2] This receptor also displayed some binding selectivity between the substrate and product, namely lysine 9 of the H3 tail and dimethyllysine 9. In ITC experiments, **A_2B** displayed 6.3 μM affinity for $K9Me_2$ and 22 μM affinity for K9 in the short H3 sequence Ac-WGGG-QTARK$(Me)_n$STG-NH_2, a 3-fold difference.[4]

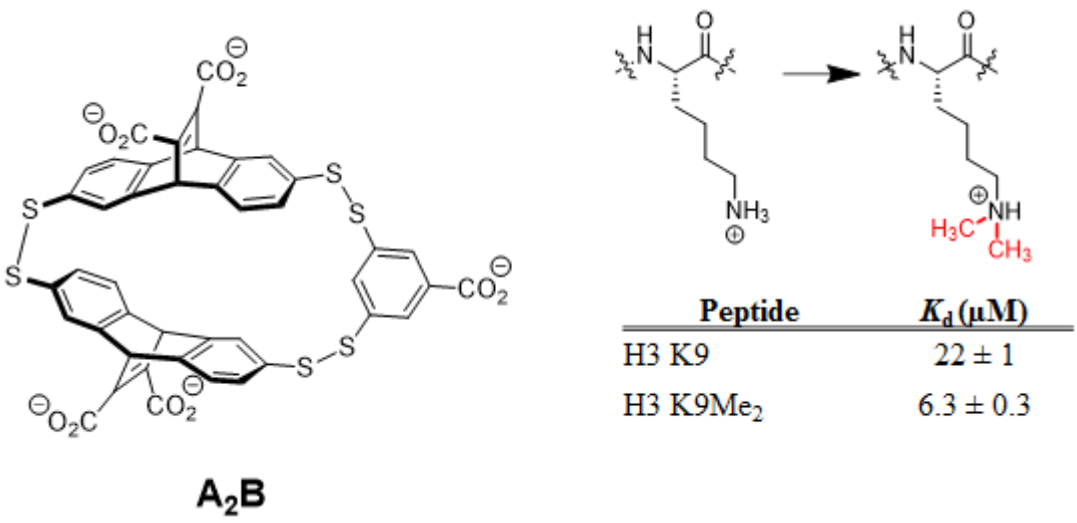


Figure 2.12 Receptor **A_2B** and binding affinities for representative peptides (Ac-WGGG-QTARKMe$_n$STG-NH$_2$) for the enzymatic methylation of lysine 9 by G9a.[4]

Because we knew that this receptor would be able to differentiate between substrate and product, we moved directly into fluorophore screening. Because an IDA relies on signal modulation upon competitive binding, we required a fluorophore that not only bound to the receptor and exhibited a fluorescence change, but also which had a strong binding affinity. We began our screens with several previously reported dyes, either commercially available or synthesized in few steps, shown in Figure 2.13.[45,46]

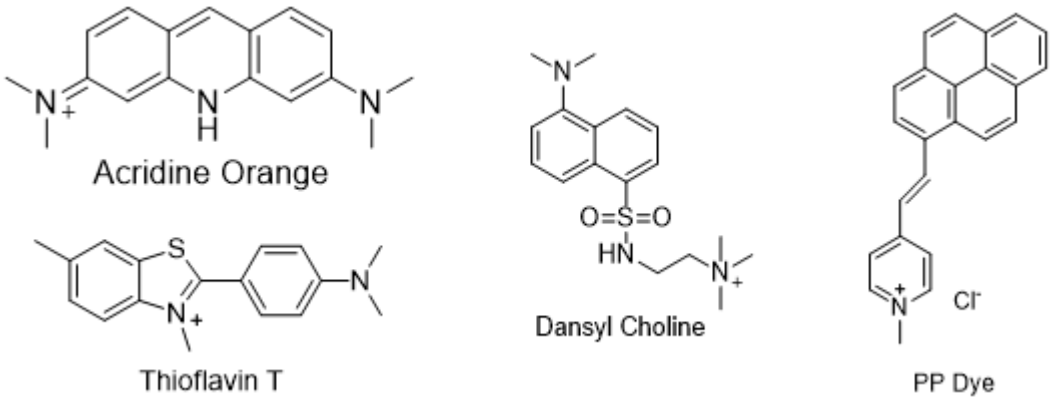


Figure 2.13 Environmentally sensitive fluorophores screened in the initial system study. Each has been previously shown to interact with a host molecule and change fluorescent properties.

Each of the dyes chosen displayed a quaternary ammonium moiety, which we believed would allow it to bind favorably with our receptors. We initially screened Dansyl Choline using a simple fluorescence experiment, where measured the fluorescence of the dye alone and then added an equal volume aliquot of A_2B as seen in Figure 2.14.

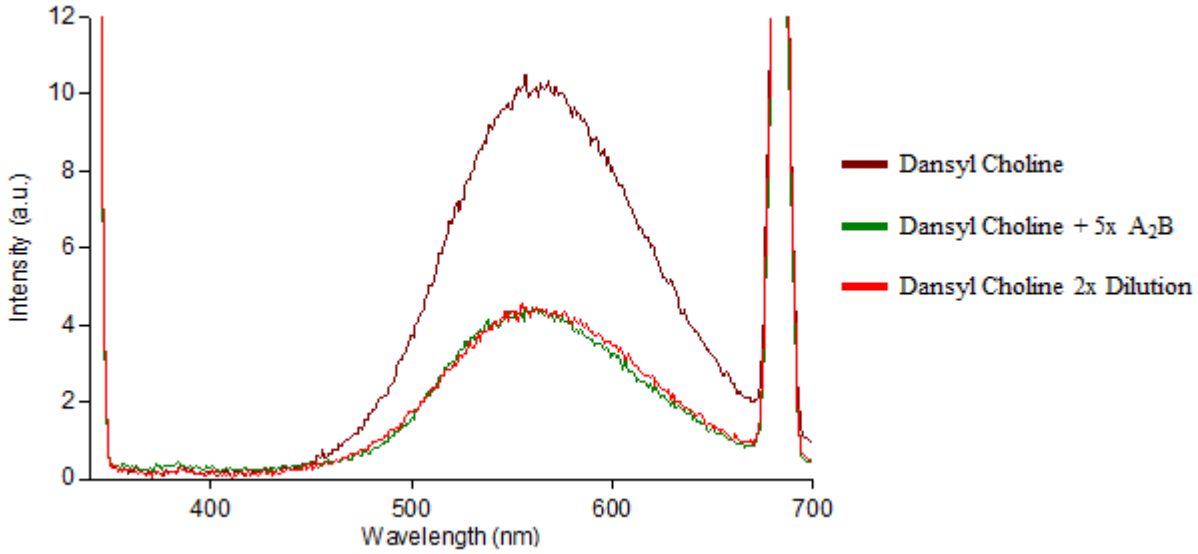


Figure 2.14 Fluorescence intensity of Dansyl Choline (58µM), Dansyl Choline (29 µM) and **A_2B** (125 µM), and Dansyl Choline (29 µM) in pH 8.5 phosphate buffer.

Dansyl choline was previously reported to have enhanced fluorescence when bound to a host system, and was reported to be relatively non-fluorescent in water.[47] While we observed a low intensity fluorescence of the dye in isolation, upon addition of **A_2B**, we observed only a further decrease in signal which corresponded to the resulting dilution of the Dansyl Choline dye. This suggested that dansyl choline either was not binding to **A_2B** or binding did not induce a change in fluorescence, prompting dansyl choline to be set aside as a potential fluorescence indicator. For the remaining dyes, we performed more thorough titration based experiments to measure the fluorescent response of the host/dye system. Titrations are advantageous because the provide insight into both binding affinities as well as concentration based fluorescence response.

The last three dyes, Thioflavin T, PP Dye, and Acridine Orange were all screened in this manner, as shown in Figure 2.15.

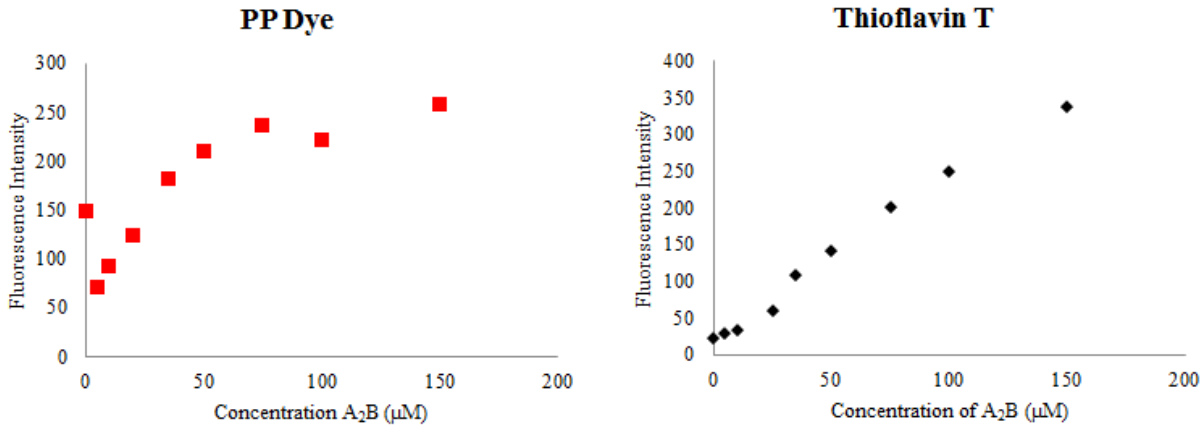


Figure 2.15 Fluorescence changes of PP Dye (5μM) and Thioflavin T (5μM) upon binding to **A_2B**.

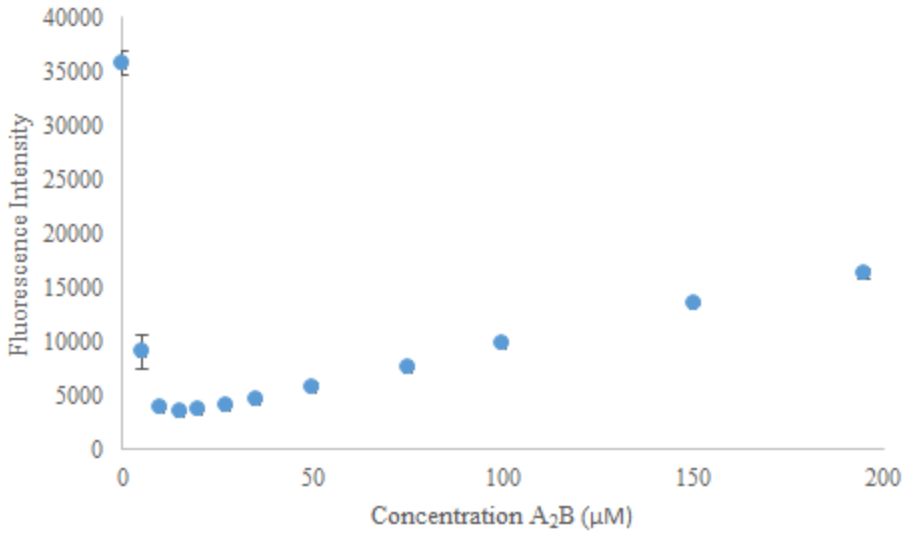


Figure 2.16 Fluorescence quenching and subsequent recovery of Acridine Orange (5 μM) with increasing concentrations of **A_2B**.

As seen from Figure 2.15, PP dye and Thioflavin T did not have ideal fluorescence properties. While they interact with the receptor and show a differential fluorescence response, both would have acted as turn off sensors, which can be more difficult to monitor. Additionally, both required almost 10-fold excess receptor to provide a 2 fold fluorescence response. Acridine orange (Figure 2.16) on the other hand displayed an interesting fluorescence response. While previous work had shown acridine orange binding to cucurbituril caused a large fluorescence enhancement,[42] our data showed an initial quench in fluorescence at approximately 1:1 and 2:1 receptor/dye pairs, then a steady increase in fluorescence at higher concentrations. We hypothesized that this observation arose from incomplete encapsulation of the AO dye by our macrocycle, and that at higher concentrations we were potentially seeing 2:1 host/guest binding. This results in a more hydrophobic environment and causes the initial fluorescence recovery compared to the quenched 1:1 binding. While acridine orange did not display the reported fluorescence properties, it still displayed a useful fluorescence response. At low concentrations of **A_2B**, the host dye complex exhibited significant fluorescence quenching. We were able to take advantage of this quenching motif by keeping our sensor concentrations at a 1:1 ratio of **A_2B** to AO, such that competitive binding of the analyte would cause a fluorescence recovery.

2.3.2 Acridine Orange/A_2B Sensor

2.3.2.1 AO/A_2B H3K9 Peptide Screen

With a receptor/dye complex established, experiments were performed to determine whether the system could differentiate between the methylated forms of lysine. With this in mind, we synthesized a variety of short H3 peptides identical to those initially studied in the **A_2B** binding experiments, mimics of the N-terminus of the H3 tail centered around Lysine 9, the site of G9a methylation. These peptides, Ac-WGGG-QTARK$(Me)_n$STG-NH_2 were then analyzed

using the optimized host/dye sensor pair, but the fluorescence data was heavily influenced by tryptophan quenching.[48] Therefore, we moved on to a tyrosine based tag, which did not influence the fluorescence of the dye. Keeping the system at a fixed concentration of 5 μM AO and 5 μM of **A_2B**, we performed titrations with the tyrosine-tagged peptides to determine the degree of difference between the unmethylated lysine substrate and the methylated lysine products, as seen in Figure 2.17.

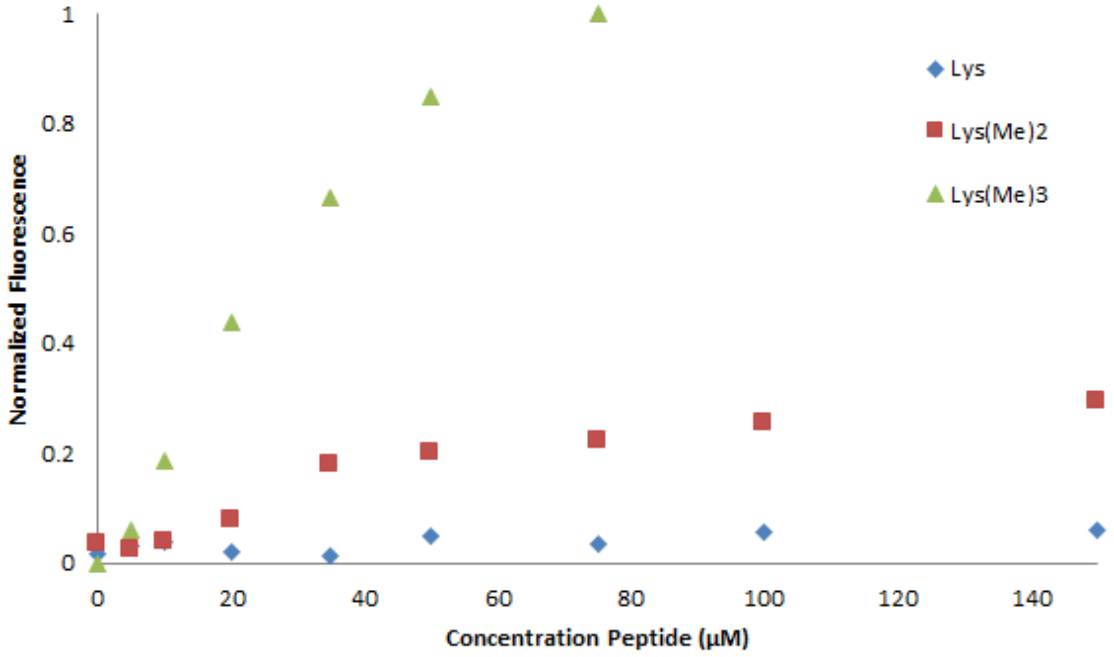


Figure 2.17 Titration of H3K9 peptides (Ac-YGGG-QTARK(Me)$_n$STG-NH_2) into the **A_2B**/AO (5/5 μM) system. Fluorescence was normalized to ensure each peptide zero concentration was also at zero fluorescence units. Experiments were run in pH 8.5 10 mM Na_2HPO_4 buffer at 27°C.

2.3.2.2 AO/A_2B Real-time G9a Assay

After performing the peptide based titrations, we began to explore real-time fluorescence assays using expressed G9a enzyme. G9a was obtained from collaborators in the Frye lab at UNC-CH, and prior to use in the **A_2B**/AO system was dialyzed out of their traditional storage

buffer, which contained dithiothreitol, a reducing agent that would degrade the receptor. We exchanged the protein into the buffer system we had used for peptide titrations, 10 mM Na_2HPO_4, pH 8.5, though 100 mM NaCl was added to ensure that the enzyme wouldn't precipitate during buffer switch. Once obtained, the enzyme was transferred into aliquots and stored in the -20°C freezer for long term use.

For the enzymatic assay, we used a 96 well plate format with the AO/$\mathbf{A_2B}$ sensor and a substrate peptide Ac-YGGG-QTARKSTG-NH_2. While this peptide is not believed to be an ideal substrate, due to loss of electrostatic contacts with the enzyme, we believed that it would be a good starting point to assess reaction viability as well as assay output. We ran the enzymatic assays in a 96 well plate to allow for multiple controls to be run simultaneously. In this case we analyzed the reaction mixture containing the sensor cocktail, the peptide substrate, SAM cofactor, and G9a enzyme over the course of fourteen hours. In addition, we ran an experiment without the peptide substrate to assess enzyme/cofactor interference and one with AO alone to verify that the fluorescence was functioning as expected over the long-term experiment.

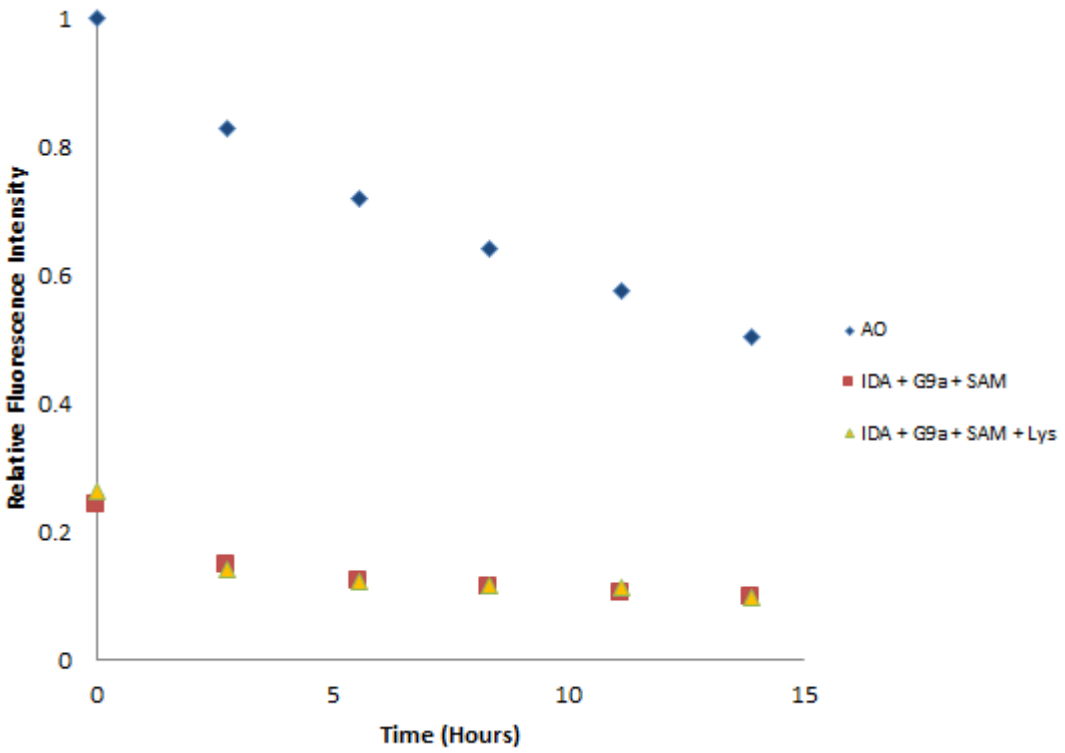


Figure 2.18 Fluorescence spectra for the enzymatic reaction of G9a on the peptide Ac-YGGG-QTARKSTG-NH_2. Two experiments contain the IDA sensor with AO/**A_2B** each at 5μM, SAM at 50 μM, and G9a at 200 nM. The enzymatic reaction also contained peptide at 15 μM. All reactions were carried out in 10 mM Na_2HPO_4 buffer, pH 8.5, with 100 mM added NaCl.

Over the time course of the experiment, we observed a steady decrease in fluorescence intensity of each reaction. If the enzymatic reaction had progressed as desired, there should have been an increase in signal as dimethyl lysine was installed, displacing acridine orange and turning on fluorescence. Instead, we observed that even for the dye alone, there was a steady decrease in signal, suggesting that there is a photobleaching process occurring. Because of this, any enzymatic activity is masked by the steady signal loss. While this fluorophore/receptor system would be well suited to a single excitation end point analysis, it will not function in real-time monitoring.

2.3.3 Lucigenin (LCG) Sensor

2.3.3.1 Receptor/LCG Binding Assay

Due to the observed photobleaching of acridine orange, we turned to the fluorophore lucigenin (LCG), shown in Figure 2.19. LCG has been proposed to undergo a photoinduced electron transfer upon complexation by an aromatic-based receptor, providing almost quantitative fluorescence quenching.[49] We hypothesized that our receptors, based around an aromatic core, would provide a similar fluorescence quenching mechanism.

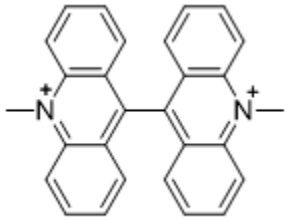


Figure 2.19. Structure of the fluorophore lucigenin (LCG).

We began by performing fluorescence titrations using a 96-well plate based method to quickly screen for affinities and quenching ability of three receptors: $\mathbf{A_2B}$, $\mathbf{A_2N}$, and $\mathbf{A_2D}$. Each receptor displays different selectivity for each methylated state of lysine, which enabled any of them to potentially function in the enzymatic assay. Because of this, each receptor went through the initial dye and short peptide screening to ensure the best receptor pair was selected for. The receptor/dye titrations are shown below in Figure 2.20.

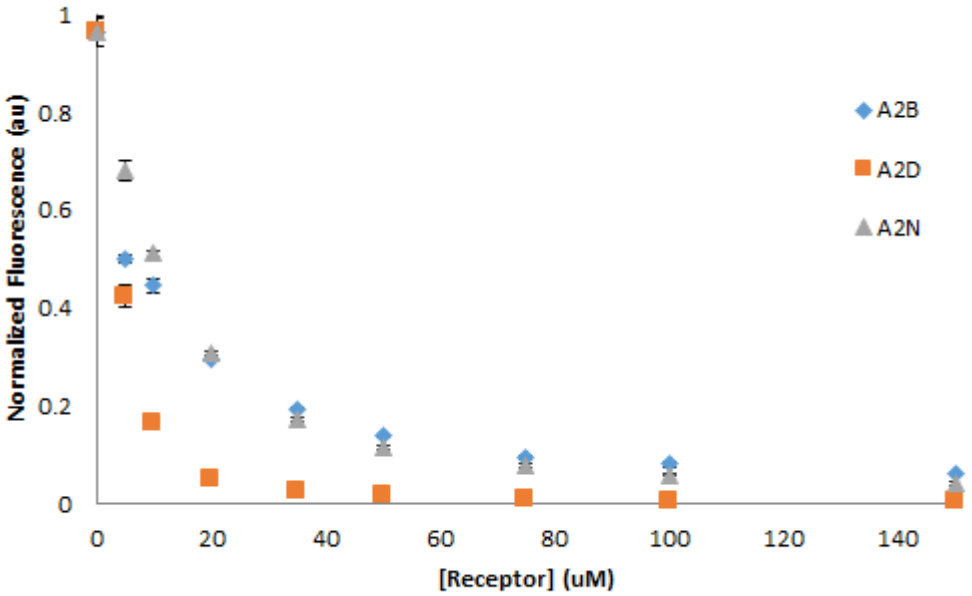


Figure 2.20 Fluorescence quenching curves for 5 μM LCG with each receptor (0-150 μM). Fluorescence signal was normalized for each receptor run, with the initial fluorescence for each experiment set to 1 au and the lowest fluorescence set to 0 au. Each experiment was run 25 mM K_2HPO_4, pH 8.0, 2mM $MgCl_2$, 1mM EDTA, and 0.01% Triton X-100.

Each fluorescence quenching experiment was curve-fit using Kalediagraph software to a fluorescence quenching binding equation to determine the K_d, shown below in Table 2.2.

Table 2.2. Binding data for **A_2B**, **A_2D**, and **A_2N** with Lucigenin (5 μM)

Receptor	**K_d (μM)**
A_2B	6.7 ± 0.7
A_2D	3.4 ± 0.5
A_2N	11.7 ± 0.8

Each receptor bound to LCG with low micromolar affinity. This is advantageous for the displacement assay system, as the binding of $K9(Me)_2$ to each receptor is approximately the same low micromolar affinity, while the binding to unmethylated lysine is much weaker,

suggesting that any receptor/LCG sensor would be able to differentiate between substrate and product of the enzymatic reaction.

2.3.3.2 Receptor/LCG Peptide Fluorescence Titrations

With the receptor dye pairs established, we transitioned into peptide displacement experiments to test the applicability of each sensor to the G9a enzymatic assay. Photobleaching experiments were performed prior to the peptide titration to ensure that the previous assay issue would not occur here, and no quenching was observed over the lifetime of the experiment. We utilized a similar 96 well plate format to rapidly screen the H3 mimic peptides, Ac-YGGG-QTARK(Me)$_n$STG-NH$_2$ with each receptor dye pair at fixed concentrations, shown in Figure 2.21 - Figure 2.23.

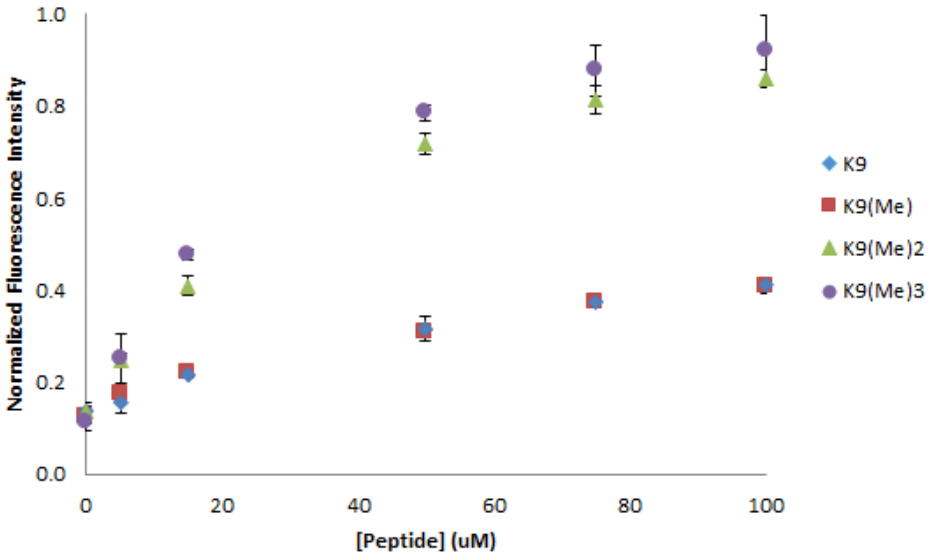


Figure 2.21 IDA for the H3 mimic peptide Ac-YGGG-QTARK(Me)$_n$STG-NH$_2$ using **A$_2$B** (10 μM) and LCG (2.5 μM). The fluorescence was normalized such that maximum fluorescence of the assay was set to one while the quenched well was set to zero.

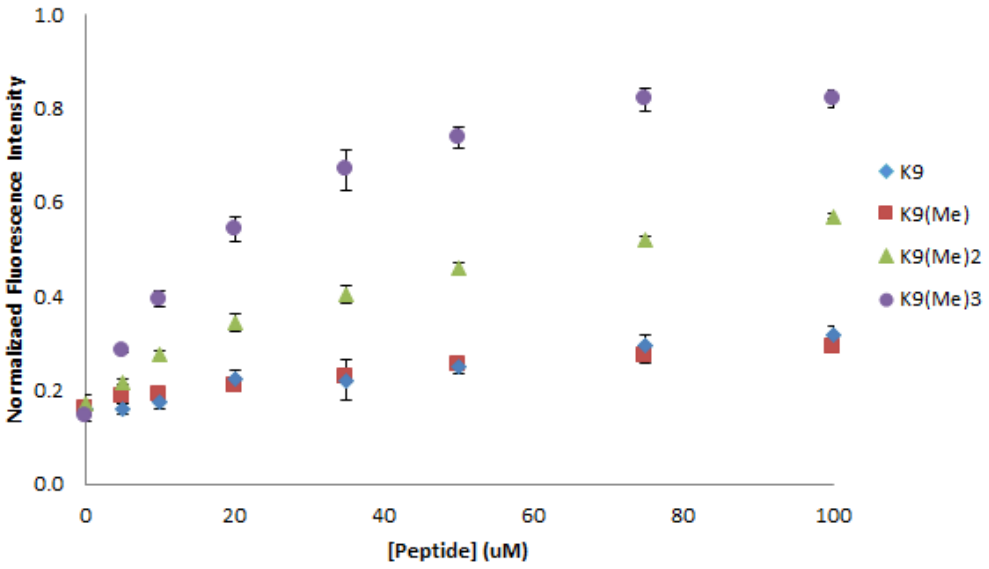


Figure 2.22 IDA for the H3 mimic peptide Ac-YGGG-QTARK(Me)$_n$STG-NH$_2$ using **A$_2$D** (5 μM) and LCG (1 μM). The fluorescence was normalized such that maximum fluorescence of the assay was set to one while the quenched well was set to zero.

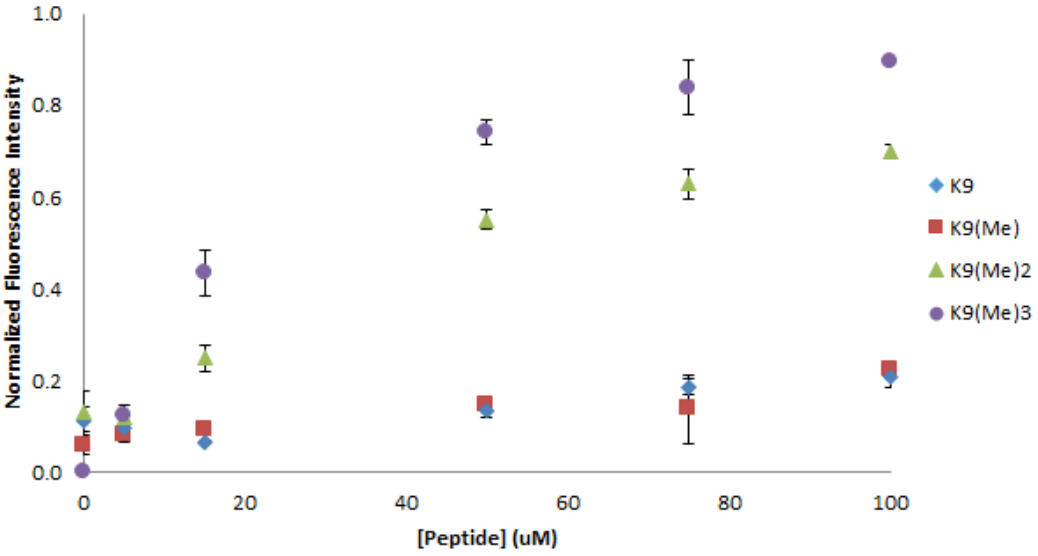


Figure 2.23 IDA for the H3 mimic peptide Ac-YGGG-QTARK(Me)$_n$STG-NH$_2$ using **A$_2$N** (15 μM) and LCG (2.5 μM). The fluorescence was normalized such that maximum fluorescence of the assay was set to one while the quenched well was set to zero.

The three receptor/dye pairs gave a different pattern of responses for each of the methylated forms of lysine. Importantly, we observed two receptors that were able to distinguish

between the mimic substrate peptide, unmethylated lysine, and the dimethylated product. In taking a closer look at **A_2B** and **A_2N** in Figures 1.21 and 1.23, we observed that **A_2N** has a lower initial response to the unmethylated substrate peptide. While that is interesting based on the tighter binding affinity of **A_2N** for the $K9(Me)_0$ peptide compared to A_2B^4, it is highly probable that this signal arose due to the nature of the competitive displacement, where dye binding can also influence assay response, not just analyte binding.

2.3.3.3 G9a Buffer Screening

While the LCG system was working well to differentiate between substrate and product, we were aware of literature precedenct for the quenching of LCG by chloride salts in solution.[50] While all efforts had previously used a potassium phosphate buffer with 2 mM $MgCl_2$, the peptide titrations gave evidence that this level of salt would not pose a large problem in differentiating between species. However, the enzyme assay used 150 mM NaCl, believed to be critical for the enzyme's storage and function. At these high salt concentrations, the fluorophore would be quenched, rendering the assay invalid. Because of this, our first goal was to screen different buffer conditions for enzymatic activity, to ensure that the fluorescence assay conditions would not inhibit enzyme activity and vice versa. In order to rapidly screen enzymatic conditions, we turned to radioactive SAM assays, shown in Figure 2.24, with detailed procedure in the SI.

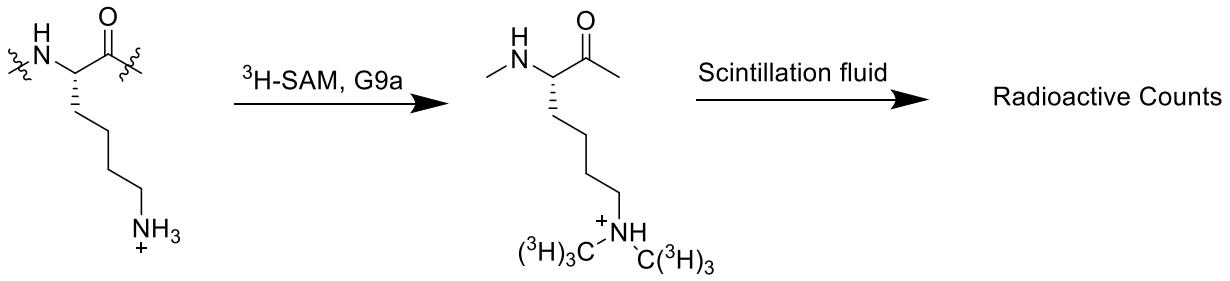


Figure 2.24 Radioactive S-Adenosylmethionine assay for the methylation of Lysine. Tritium-labeled SAM is incubated in place of the traditional SAM cofactor prior to the enzyme assay. After the reaction is complete, methylated peptide is isolated and the radioactivity is counted, providing relative amount of methylation incorporation.

The radioactive assay allowed us to rapidly examine several different buffers for activity. Each set of experiments was run using the first 20 amino acids of the H3 tail, a commonly accepted substrate, with the H3 1-20 K9$(Me)_3$ peptide as a negative control. Because of the principles of the experiment, it results in a qualitative readout, rather than the absolute amount of methylation. Therefore, each set of experiments was numerically comparable only to others in the same set , though trends were still seen in multiple runs. The first set of tests included two primary buffer systems, the potassium phosphate buffer used in previous peptide displacements as well as a glycine buffer used in the Nau Dim-5 system, though at pH 9.15 instead of pH 10. The specific buffer and buffer additives are shown below in Table 2.3, with the radioactive assay results in Figure 2.25.

Table 2.3 Buffer identities for the radioactive methylation of H3 1-20 by G9a

Trial	Buffer Type	[Buffer] (mM)	pH	[NaCl] (mM)	[$MgCl_2$] (mM)	[EDTA] (mM)	% Detergent
1	Tris	50	9	100	-	-	-
2	K Phos	25	8	-	2	1	0.01
3	K Phos	25	8	-	1	1	0.01
4	K Phos	25	8	-	0	1	0.01
5	K Phos	25	8	-	2	0	0.01
6	K Phos	25	8	-	2	1	-
7	Glycine	50	9	-	-	-	-
8	Glycine	15	9	-	-	-	-
9	Glycine	50	9	-	-	-	0.01

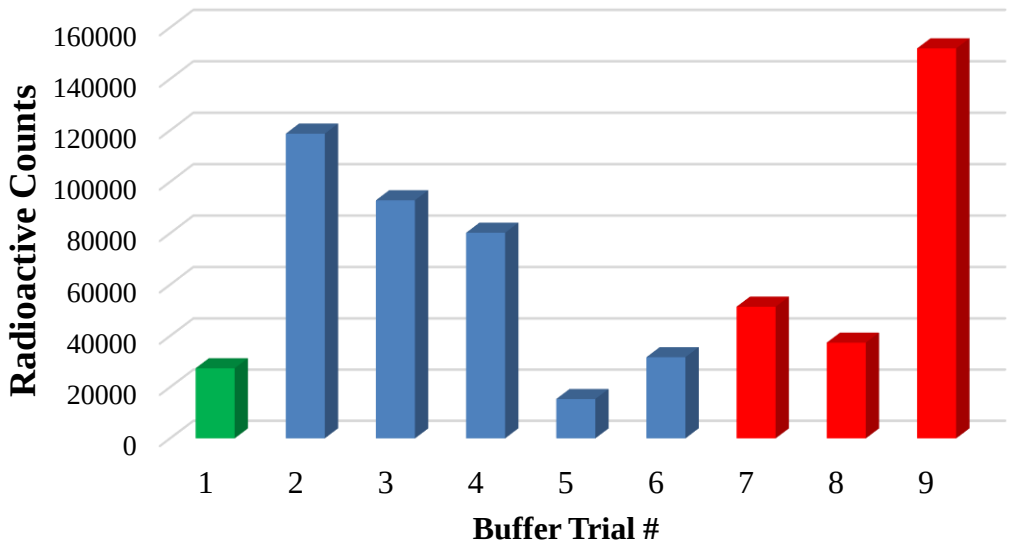


Figure 2.25 Radioactivity assay results for the methylation of H3 1-20 by G9a and ^{3}H-SAM. Buffer trial # correlates to Table 2. The green trial represents Tris control buffer, blue represents potassium phosphate buffer with various additives, and red represents varying degrees of glycine buffer.

The traditional buffer of choice for performing the G9a assay, which was 50 mM Tris buffer, pH 9, with 100 mM NaCl, was actually the worst buffer in our experiments. This result

aligns with previous studies, which showed an inhibitory effect of high salt concentrations on G9a.[51] In the radioactive test the phosphate based buffer we used for initial displacement tests, buffer 2, gave a very strong result, over 5x more effective than the traditional Tris buffer. Lowering the salt concentration saw activity begin to drop off, however, removal of the EDTA at the same concentration of salt saw activity almost halt altogether. This suggested that in isolation, the lack of either component could be acting in an inhibitory role. Buffer 4 however, lacking $MgCl_2$, would be a strong alternative in our assay, because while there is a slight activity trade-off it lacks all sources of Cl^-, removing its ability to quench the fluorophore.

With glycine buffer, both at 50 mM and 15 mM resulted an increase in activity over the Tris control, but not to the same levels as the phosphate buffers. However, upon addition of the Triton X-100 detergent at 0.01% by volume (Buffer experiment 9, Figure 2.25), we observed the highest levels of methylation in the assay. When triton was excluded from the phosphate buffers, activity dropped to around the same level as the original glycine buffers, still above the control, but now between 1.5x and 2x activity. We believe that the detergent is helping to maintain G9a in the reaction medium. The assay is carried out in a plastic 1.5 mL eppendorf tube, and it is possible that without detergent the protein can stick to the walls of the reactor, minimizing reactivity, which is mitigated by the addition of the detergent.

2.3.3.4 G9a Substrate Screening

Our next step was to optimize the H3 substrate. Our chosen sensor system utilized **A_2N**, which has six carboxylates decorating the rim for water solubility, but they can also interact non-specifically with peptides bearing multiple basic residues.[4] The H3 tail used in the above radioactive assays was H3 1-20, with a free N-terminus, giving an overall peptide charge of +8. This high positive charge could potentially abolish the selectivity of the system, if there is high

binding affinity regardless of methylation. Because of this, we synthesized two other substrates, H3 1-15 and H3 1-12, both bearing a free N-terminus for net charges of +6 and +5 respectively. We studied these six peptides, all three lengths bearing K9 and K9(Me)$_2$ using our **A$_2$N**/LCG sensor system.

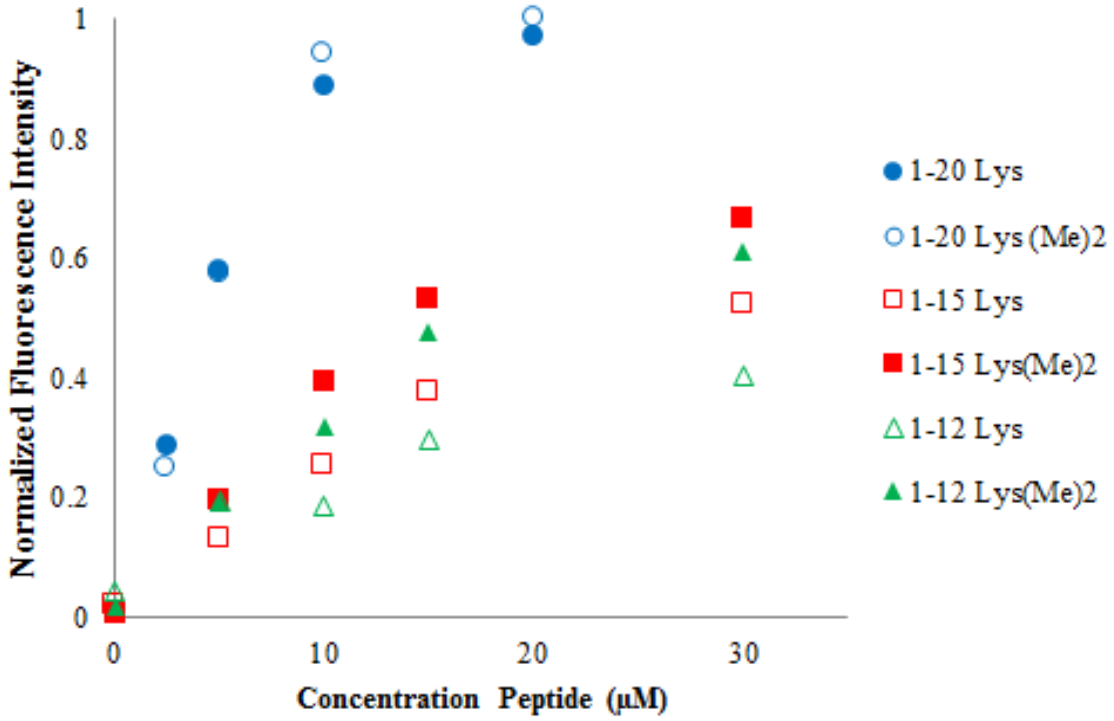


Figure 2.26 Peptide displacement tests to determine optimal substrate length using the **A$_2$N** (25 µM)/LCG (2.5 µM) in 25 mM K_2HPO_4, 2mM $MgCl_2$, 1mM EDTA, 0.01% Triton X-100. Each peptide was examined in both the methylated and unmethylated state.

In the resulting fluorescence assay, the predicted outcome with H3 1-20 was observed, namely that the unmethylated substrate displaces the fluorophore at the same level as the methylated product peptide. This proves a large problem, because even if the enzyme assay would go 100%, there would only be approximately a 6% signal difference, reducing the utility of the system. There was good signal difference in the H3 1-15 and H3 1-12 peptides, with net charges of +6 and +5 respectively, however, suggesting that these would be good targets for the

fluorescence assay without sacrificing enzymatic activity. To test this, we performed another round of radioactivity tests with the H3 1-15 substrate, to compare activities and ensure that activity did not diminish significantly. We observed that the H3 1-15 peptide had approximately 55% enzymatic activity compared to the H3 1-20 peptide, both using the glycine buffer system. While the loss of activity was not ideal, it was the best balance between enzymatic activity and assay response.

2.3.3.5 Fluorescence Assay Response Optimization

Based on the results described above, we moved forward with the H3 1-15 substrate peptide and the 50 mM glycine buffer, pH 9.15. H3 1-15 afforded a strong balance of enzymatic activity as well as fluorescence activity, and the glycine buffer removes the potential for chloride quenching. With substrate and buffer established, the fluorescence assay needed tuning to provide a maximal response. In the fluorescence assay, we could optimize three main component concentrations, the fluorophore, the sensor, and the substrate/product peptide. Because of the sheer number of possible combinations, we turned to a high-throughput, 384 well plate based approach. To do this, we set up several plates, each with a different fixed LCG concentrations, then varied the amount of both receptor and paired substrate/product peptide. By subtracting the difference in signal between the product displacement and substrate displacement, we could get a good picture of the assay response at many different concentrations in one experiment.

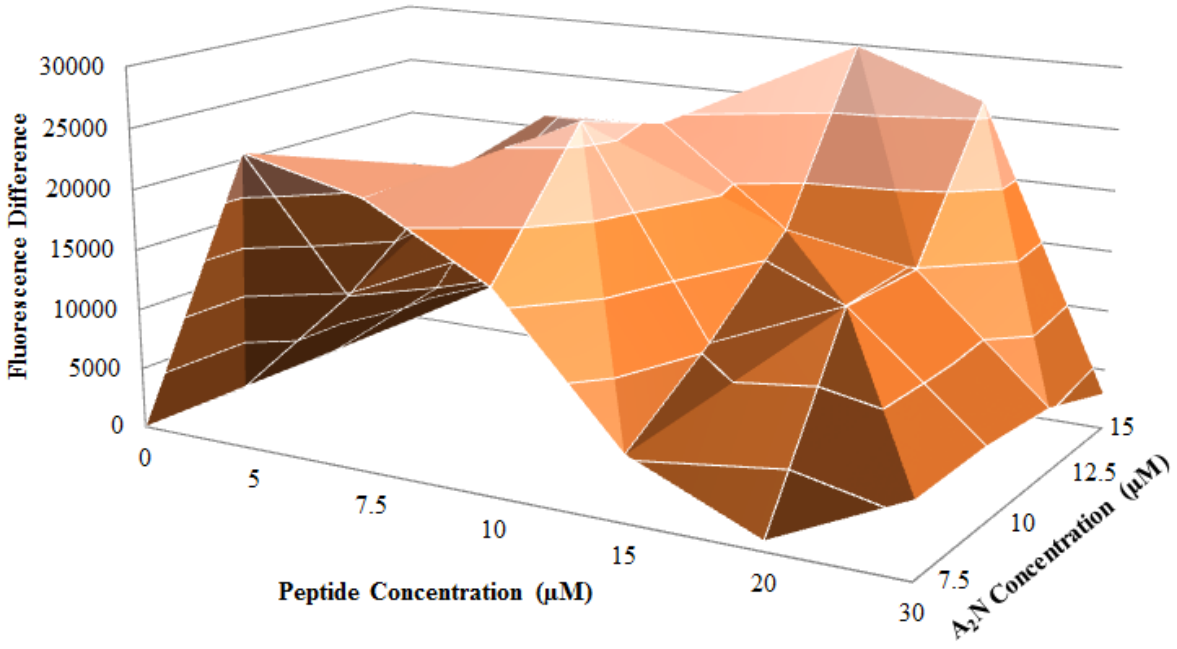


Figure 2.27 Fluorescent 'heat map' of **A_2N**/LCG (1 μM) sensor system for H3 1-15 (ARTKQTARKSTGGKAY-NH_2). The fluorescence difference was calculated by subtracting the response of H3 1-15 from H3 1-15 K9$(Me)_2$, therefore higher difference is better signal differentiation between the two species.

In the resulting heat maps, a LCG concentration of 1 μM gave the best assay response, so it was the focus of our next optimization efforts. As evidenced from Figure 2.27, the heat map provided a good region of strong signal difference. We noticed a sharp falloff of signal at higher peptide concentrations, the unmethylated peptide displaced the dye to such an extent that methylation itself played no role. At approximately equimolar concentrations of peptide to receptor we observed the best signal. While in each case there is some displacement by the unmethylated peptide species, the sensor is still able to preferentially sense the methylated species.

The objective of the heat map was to quickly identify regions of concentrations that would provide good signal differentiation. However, the planned assay format was using a

cuvette based reader instead of the microplate format, and the format switch can preclude direct translation of the results. Because of this, we explored several of the successful hits from the plate-based assay using a fluorescence cuvette. In these experiments, a starting volume of LCG was added to the cuvette, then much smaller volumes of high concentration component were added to prevent significant dilution. In each case, **A_2N** was added to the cuvette, followed by H3 1-15 and finally an equal concentration of H3 1-15 K9$(Me)_2$, allowing us to confirm the fluorescence differentiation. One such experiment is shown below in Figure 2.28.

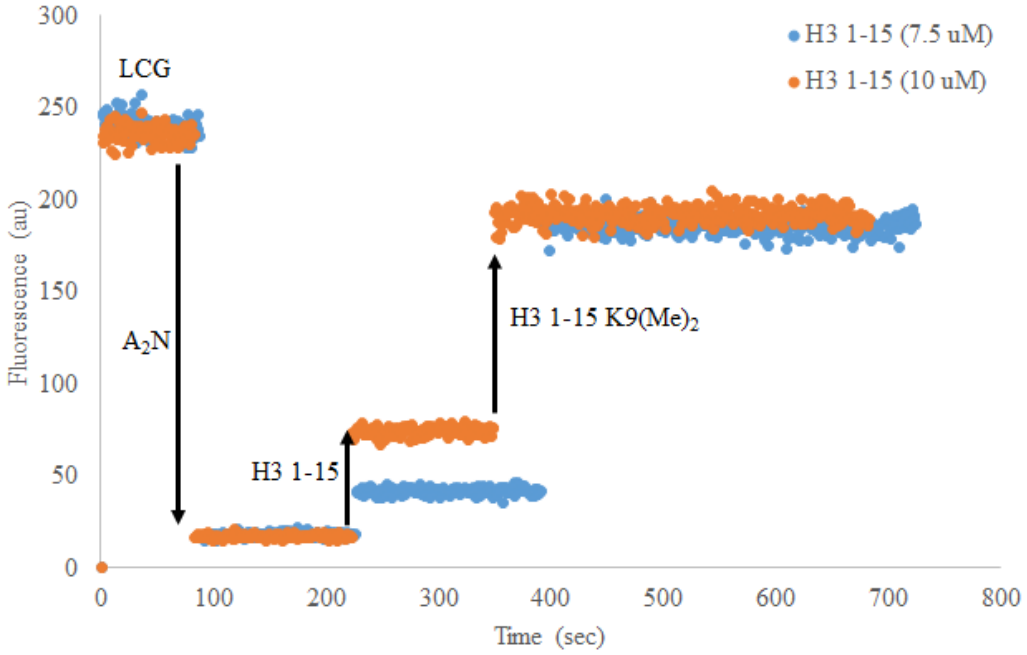


Figure 2.28 Cuvette based fluorescence differentiation assay. The sensor pair, **A_2N** (7.5 μM) and LCG (1 μM), were titrated with varying concentrations of H3 1-15 peptides either unmodified or bearing K9$(Me)_2$. The experiments were performed in 50 mM glycine, pH 9.15 at 25°C.

The cuvette experiments were able to identify several sensor/peptide concentrations that gave strong difference between substrate background and product signal. In each case, addition of peptide substrate caused a moderate recovery of fluorescence, though the signal jump was

much higher with the dimethylated H3 1-15. We did observe however that in no situation did the fluorescence recover completely, which while not optimal, should not negatively impact the enzymatic assay.

2.3.3.6 Real-time Fluorescent G9a Assay

Using the identified sensor conditions, we began to examine enzymatic assays using active G9a and SAM cofactor. To do this, we performed cuvette assays using a similar method to the fluorescence displacement experiments previously described. The first experiments ran were controls, reaction cocktails missing either G9a or SAM cofactor to ensure there wasn't any fluorophore displacement, and are shown in Figure 2.29.

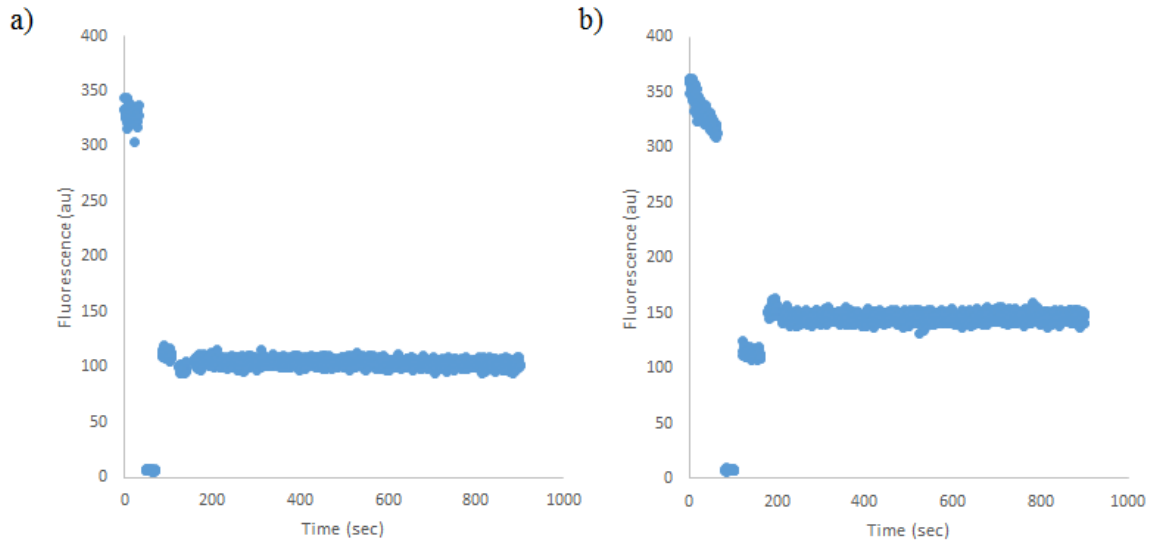


Figure 2.29 Cuvette based fluorescence enzyme assay. The sensor pair, **A_2N** (12.5 µM) and LCG (1 µM), were titrated with unmodified H3 1-15 peptide substrate (10 µM) in 50 mM Glycine buffer, pH 9.15. a) +SAM cofactor control (50 µM) b) +G9a enzyme control (1 µM)

In the presence of both G9a and the SAM cofactor, we observed no additional displacement of fluorophore beyond the expected increase due to the H3 1-15 substrate background. This confirms that the cationic S-Adenosylmethionine does not interact with the

sensor system and causing fluorescence recovery. Additionally, it shows no interaction from G9a, suggesting that not only does the enzyme itself not interact with the sensor, but also the buffer dialysis to remove DTT was successful, as reducing agent would cause a fluorescence increase over time as the receptor degraded. Following these controls, we explored the assay's capability to monitor the enzymatic methylation of H3 1-15 by G9a in real time.

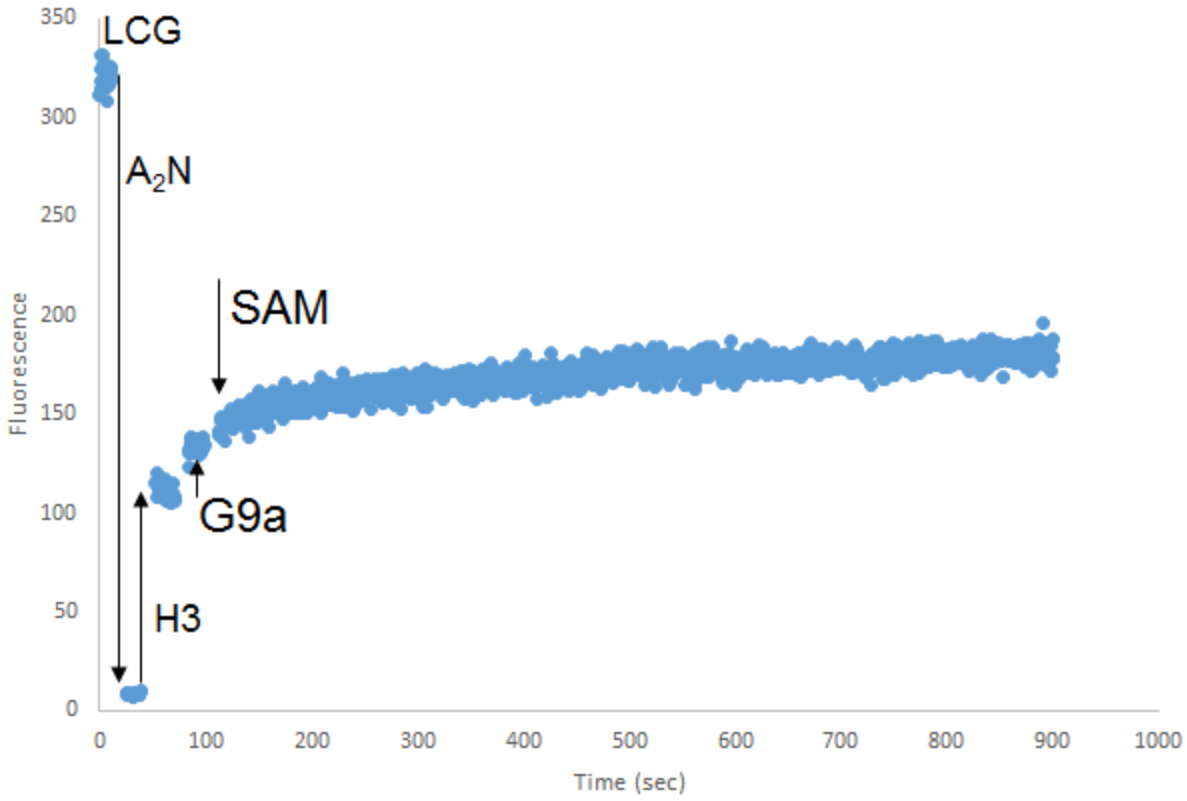


Figure 2.30 Real-time G9a fluorescence assay, 50 mM glycine, pH 9.15, 37°C. The arrows indicate the readout was paused while the components were injected into the reaction mixture. LCG (1 μM) was quenched by **A_2N** (12.5 μM), followed by background fluorescence recovery by the H3 1-15 peptide substrate (10 μM). Addition of G9a (1 μM) caused a slight fluorescence increase, and addition of SAM (50 μM) initiates the reaction, which is monitored for approximately 10 minutes.

After addition of SAM, there was a steady increase in fluorescence, suggesting that the enzymatic reaction was installing the dimethylation mark on lysine, thereby increasing the affinity of the peptide and displacing the dye from the receptor. However, the initial velocity of

the experiment was not very fast, in fact over the entire time course of the run there was only an 11% fluorescence difference between start point and end point, compared to the 50% fluorescence difference in the initial cuvette studies. We repeated the experiments several times with varying conditions, one of which is shown in Figure 2.31, though never saw an increase above this 10% fluorescence recovery,

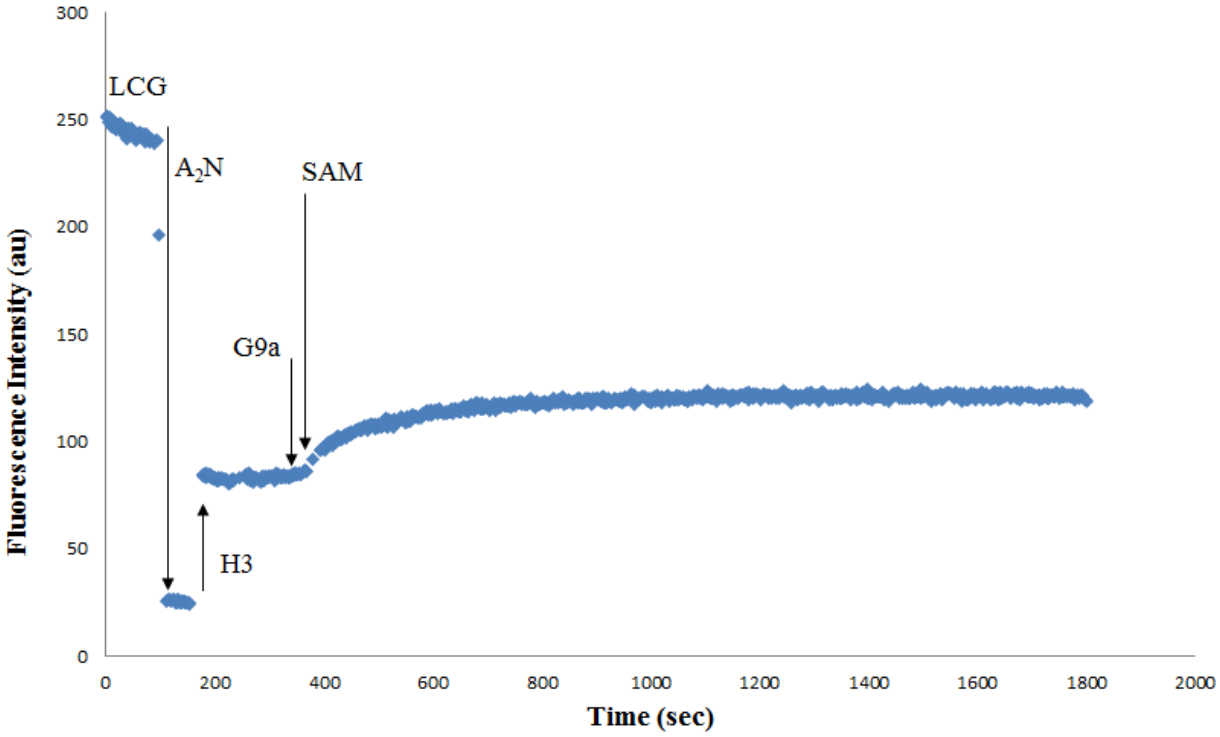


Figure 2.31 Real-time G9a fluorescence assay, 50 mM glycine, pH 9.15, 37°C. LCG (1 μM), **A_2N** (15 μM), H3 1-15 (15 μM), G9a (0.2 μM), and SAM (300 μM).

While the assay is in fact working, the 10% change in fluorescence does not provide a large enough window to do the experiments we hoped for, namely enzymatic parameter characterization and more complex peptides bearing multiple modifications. Since the fluorescence response is so small, minor perturbations in activity from things such as a neighboring PTM would most likely fall within error of the original peptide. We believe this low response could be due to several factors. While we verified the enzymatic activity of the shorter

peptide substrates using the radioactive assay, we have no way of actually quantifying the amount of methylation incorporated into the substrate. So while active, the enzymatic reactions may not be proceeding to completion, minimizing the potential fluorescence response of the assay. Likewise, while we see background fluorescence displacement from the substrate peptide, the receptor could also be acting in an inhibitory manner to the assay, preventing methylation from occurring rapidly and prematurely shutting down the methylation, as evidenced by the leveling off of the fluorescence signal in Figure 2.31. Unfortunately, if the concentration of receptor is lowered while substrate remains constant, the background signal increases dramatically, further lowering the fluorescent assay response.

Application of the **A_2N**/LCG sensor to the real time monitoring of G9a methylation was successful, though the assay conditions and enzyme conditions still require further optimization. In the context of the **A_2N**/LCG sensor, the drop off in initial velocity without full fluorescence recovery suggests that the reaction is stalling. Of the three receptors studied, **A_2N** is reported with the tightest affinity for unmethylated lysine, so while it did display the best signal in the short peptide assays shown in Figure 2.23, it might not be amenable to the longer peptide substrates. Additionally, switching the peptide substrate back to the shorter H3 mimic, YGGG-QTARKSTG-NH_2 should still be active in the assay but not be inhibited by receptor binding, allowing for longer reaction times to monitor activity while still providing a useful platform for studying combinatorial modifications and enzyme inhibitor assays. With several avenues of further optimization to pursue, the application of DCC based receptors to an enzymatic IDA is a promising approach to studying the lysine methyltransferase reaction.

2.4 Arginine Methyltransferase Assay

2.4.1 PRMT Sensor Design

While developing the indicator displacement assay for lysine methylation, we also explored its utility in monitoring other SAM dependent methyltransferases. The type-1 PRMTs, responsible for installing the asymmetric dimethylation of arginine in histone tails. The previous fluorophore studies had shown that LCG bound to **A_2D** and could differentiate lysine methylation. However, **A_2D** was originally designed as a receptor for asymmetric dimethyl arginine, with binding affinities shown below in Table 2.4.[3]

Table 2.4 Binding affinities of **A_2D** for the H3 tail peptide Ac-YGG-QTAR$(Me)_n$STG-NH_2 in 10 mM Na_2HPO_4 pH 8.0. Dimethyl arginine is shown in one of two forms, symmetric (*s*RMe_2) or asymmetric (*a*RMe_2).

Peptide	K_d (μM)
H3 R	≥ 60
H3 R(Me)	26.0 ± 3
H3 *s*R$(Me)_2$	38.4 ± 4.8
H3 *a*R$(Me)_2$	5.1 ± 0.6

Based on the binding data above, we hypothesized that the **A_2D**/LCG sensor would be able to sense the methylation of arginine in a real-time fluorescence assay. We chose to study PRMT1, which installs asymmetric dimethyl arginine on R3 of the H4 tail. This enzyme has been well studied, and several inhibitors have been developed, allowing us to test our assay against currently published results to ensure that the assay is working as intended.[52,53]

2.4.2 A_2D/LCG Methylated Arginine Peptide Displacement

Similar to the G9a methylation assay, several optimization steps had to be performed. As in the case of G9a, the most commonly presented substrate for PRMT1 is H4 1-20, however, that sequence of the histone tail is highly positively charged, notably residues 15-20, which are a dense patch of charge, KRHRK.[53] Because of this we believed that the H4 1-20 sequence would

simply displace the fluorophore itself, rendering the assay insensitive to methylation. Therefore, we shortened the peptide sequence and ran an initial peptide displacement titration, comparing the H4 1-15 sequence to the H4 1-8 sequences, as shown in Figure 2.32.

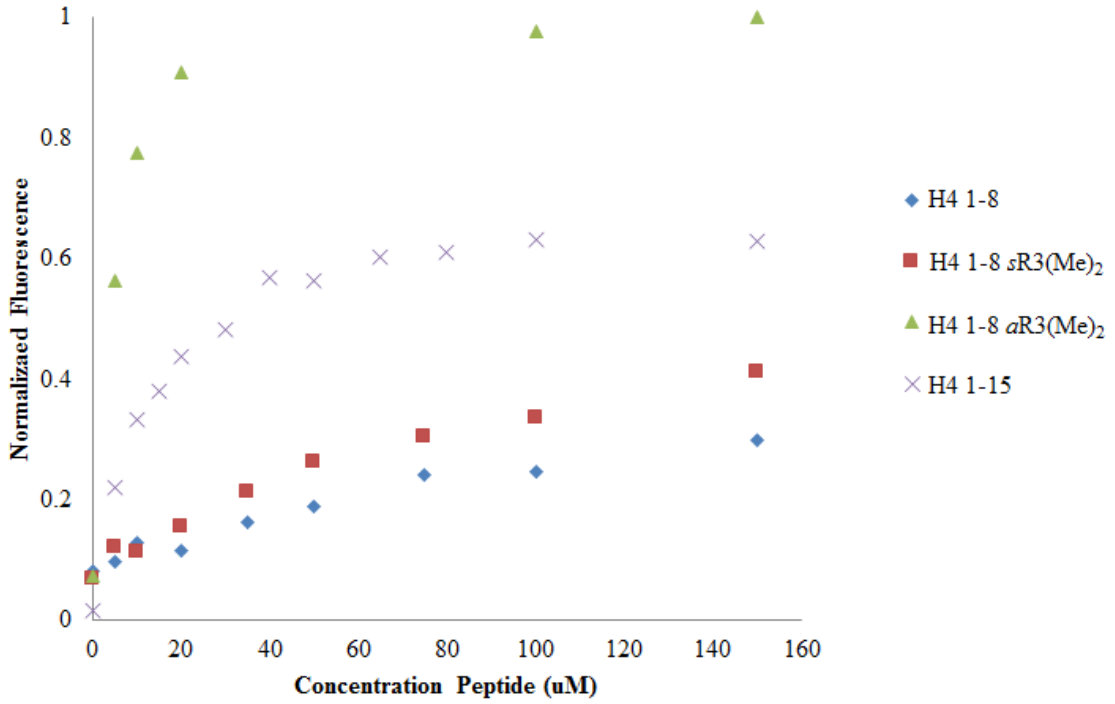


Figure 2.32 Fluorescence titration of H4 peptides into the **A_2D**/LCG sensor system (5 μM/1 μM). H4 1-8 (Ac-SGR(Me)$_n$GKGG-GY-NH_2) and H4 1-15 (Ac-SGRGKGGKGLGKGGAY-NH_2) were all analyzed in 25 mM K_2HPO_4, pH 8.0, 2mM $MgCl_2$, 1mM EDTA, 0.001% Triton X-100. Fluorescence was normalized to ensure 0 μM peptide was at 0 au and the maximum observed fluorescence was set to 1 au.

The H4 1-15 peptide significantly displaced LCG from **A_2D**, however, the asymmetrically methylated H4 1-8 still had a much higher signal. This is beneficial, because even though the 1-15 substrate has an extra +2 positive charge, we still observe differentiation, suggesting that if H4 1-15 were dimethylated the assay response would be strong.

2.4.3 PRMT1 Radioactive Activity Assay

To test enzymatic activity, the PRMT1 enzyme was secured from a collaboration with Brian Strahl's lab at UNC-CH. Working with them, we expressed GST-tagged rat PRTM1 and tested it for activity in the radioactive assays, alongside a human PRMT6 provided by the Strahl lab, which methylates R2 on the H3 tail. However, in the case of both enzymes we observed baseline radioactivity, suggesting that the enzymatic reaction was not taking place. While we expected the H4 1-15 substrate to have a lower activity based on previous reports,[53] it should have still reacted in the radioactive enzymatic assay, suggesting that the expression did not produce active enzyme. Likewise, PRMT6 was run in the presence of H3 1-20, a natural substrate, though it too did not display activity. Currently work is being done by Lauren St. Louis in our lab to run and optimize the PRMT1 expression to ensure that active enzyme is available. Once the enzyme activity is verified, the fluorescence assay can be optimized using procedures described above to produce a real-time label free enzyme indicator displacement assay.

2.5 Experimental

2.5.1 Peptide Synthesis

All peptide synthesis was performed on a Tetras Peptide Synthesizer using CLEAR-Amide resin from Peptides International using Fmoc N-terminal protected amino acids with protected side chain functionality. Coupling reagents were HOBt/HBTU in DMF with 8 equivalents of DIPEA.. After synthesis, all peptides were acylated using 5% acetic anhydride and 6% 2,6-lutidine in DMF, followed by cleavage and global deprotection using 95% TFA, 2.5% TIPS, and 2.5% H_2O for four hours. Trimethylated peptides were synthesized by adding 2 equivalents of Fmoc-Lys$(Me)_2$-OH. Following acylation, but prior to cleavage, the dimethyl lysine was further methylated using ten equivalents of MTBD and MeI. Peptides were purified using semi-preparative reverse phase HPLC using a XBridge Peptide C18 column with a linear

gradient of A and B (Solvent A: 95% H_2O/5% CH_3CN with 0.1% TFA; Solvent B: 95% CH_3CN/5% H_2O with 0.1% TFA) and monitored at 214 nm. Peptides were then lyophilized and characterized by ESI-MS.

2.5.2 Fluorescence quenching

Fluorescence quenching experiments were performed using purified receptors and commercially purchased Thioflavin T, Acridine Orange and Lucigenin. PP Dye was synthesized following a published procedure.[45] AO, Thioflavin T, and PP Dye were all studied using 10 mM Na_2HPO_4 buffer, pH 8.5. LCG titration used 5 μM of fluorophore and increasing concentrations of receptor in pH 8.0 25 mM K_2HPO_4, 2 mM $MgCl_2$, 1mM EDTA, and 0.01 Triton X-100. Receptor concentrations were determined using reported extinction coefficients.[2–4] Plates were centrifuged and incubated for 15 minutes prior to reading on a POLARStar Omega (BMG Labtech) using ex 485nm, em 520nm for AO and ex 370nm, em 510nm for LCG. For Thioflavin T and PP Dye, plates were read using a SpectraMax M5 instrument. The fluorescence quenching data for LCG was fit using KaleidaGraph to the following equation:[54]

$$I = \frac{\left[I_0 + I_\infty \left(\frac{L}{K_d}\right)\right]}{\left[1 + \left(\frac{L}{K_d}\right)\right]}$$

I is the observed fluorescence intensity, I_0 is the initial fluorescence intensity, I_∞ is the fluorescence intensity at binding saturation, [L] is the concentration of added receptor, and Kd is the dissociation constant.

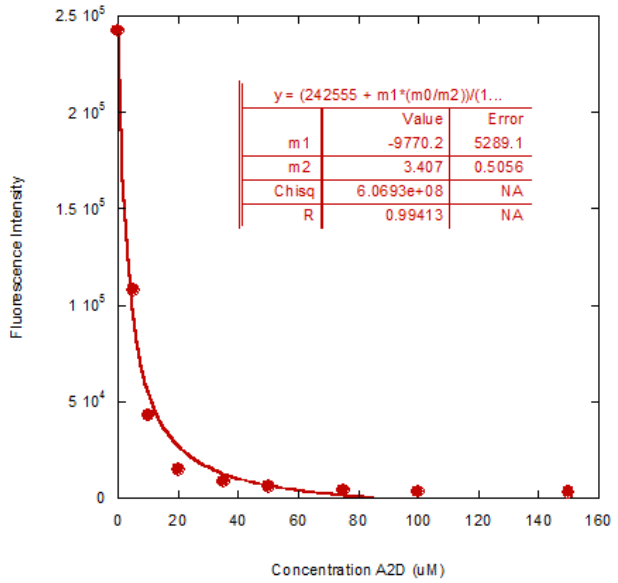


Figure 2.33 Fluorescence quenching of LCG (5 μM) binding to **A_2D** (pH 8.0 25 mM K_2HPO_4, 2mM $MgCl_2$, 1mM EDTA, 0.01% Triton X-100, 27°C)

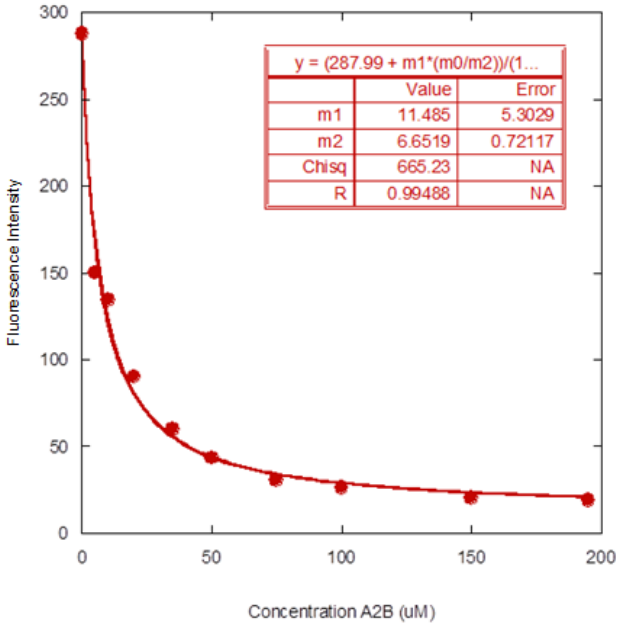
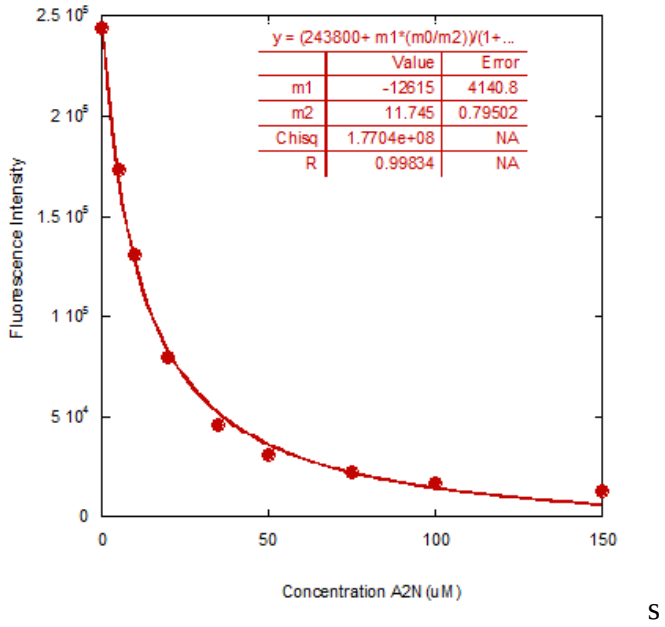


Figure 2.34 Fluorescence quenching of LCG (5 μM) binding to **A_2B** (pH 8.0 25 mM K_2HPO_4, 2mM $MgCl_2$, 1mM EDTA, 0.01% Triton X-100, 27°C)



s

Figure 2.35 Fluorescence quenching of LCG (5 μM) binding to **A_2N** (pH 8.0 25 mM K_2HPO_4, 2mM $MgCl_2$, 1mM EDTA, 0.01% Triton X-100, 27°C)

2.5.3 Peptide Fluorescence Displacement Experiments

Peptide titrations were performed in Costar 96 well half area black plates, NBS treated, or Corning 384 well half volume black NBS treated plates using the conditions described above. Each well contained the indicated concentrations of fluorophore and receptor with increasing concentrations of peptide. The fluorescence was normalized such that the highest fluorescence recovery over each of the experiments plotted was set to 1 while the lowest fluorescence signal was set to 0. Titrations were run in triplicate to ensure no major errors occurred. Peptide concentrations were determined using the extinction coefficient of Tyrosine, 1405 $cm^{-1}M^{-1}$ at 274nm.

2.5.4 Cuvette fluorescence assay

Cuvette fluorescence titrations were performed using a Cary Eclipse Fluorimeter with temperature controller. Assays were performed at a total assay volume of 95μL. The experiment

used 80µL of LCG in the cuvette and fluorescence was monitored using ex 370, em 510, PMT detector on high. The experiment was paused approximately every 100 seconds in order to add the subsequent component, in the order receptor, unmethylated lysine peptide, and methylated lysine peptide, all in 5µL aliquots. The signal of unmethylated peptide was subtracted from methylated peptide to determine the fluorescence difference between signal and background.

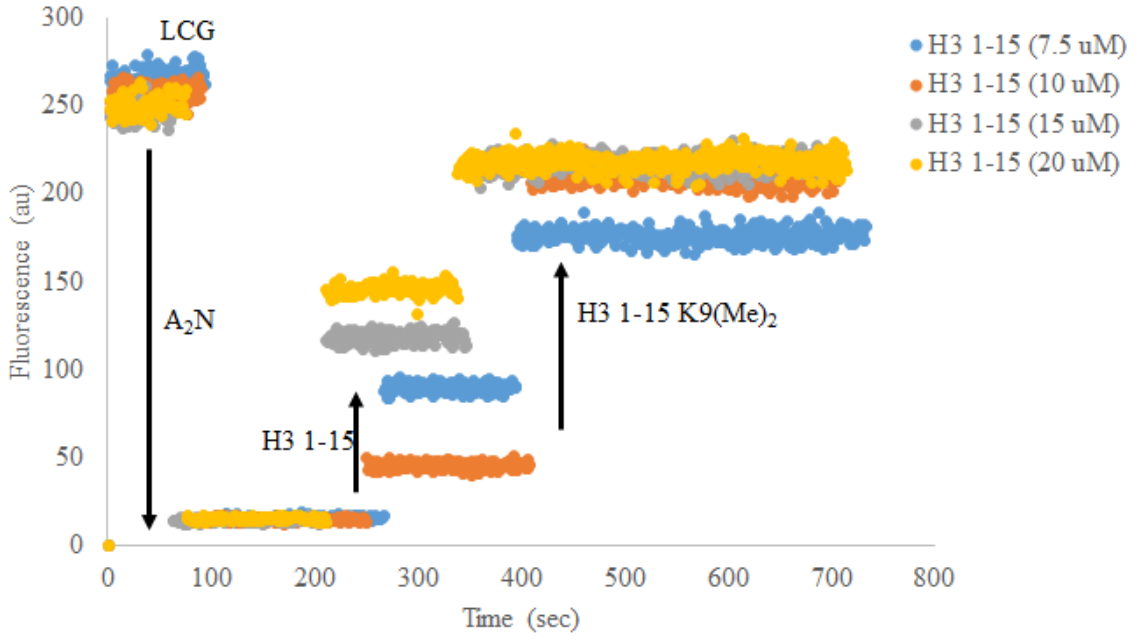


Figure 2.36 Cuvette based fluorescence differentiation assay. **A_2N** (10 µM) and LCG (1 µM), titrated with H3 1-15 peptides either unmodified or bearing $K9(Me)_2$. The experiments were performed in 50 mM glycine, pH 9.15 at 25°C.

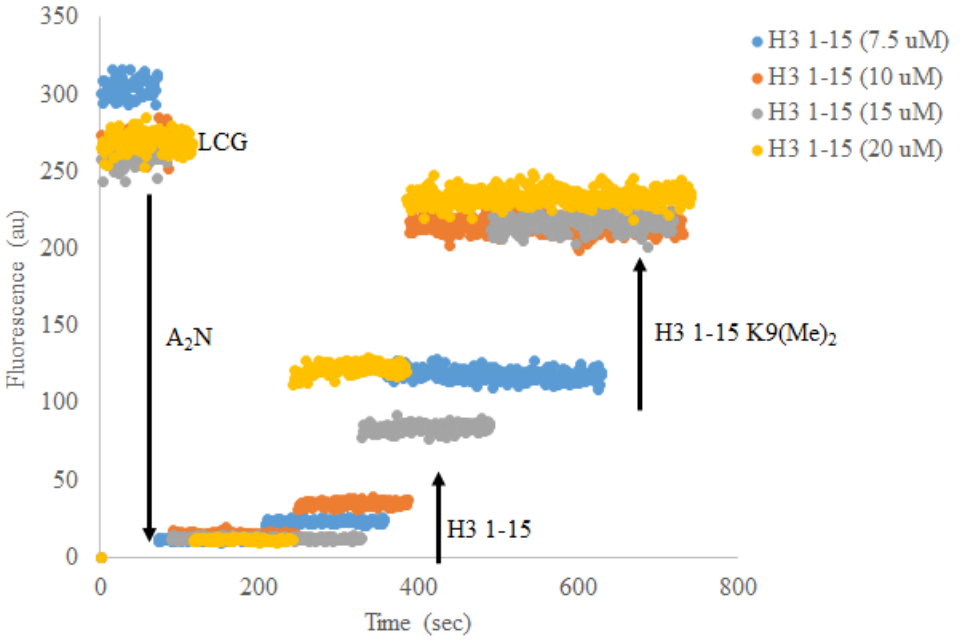


Figure 2.37 Cuvette based fluorescence differentiation assay. **A_2N** (12.5 µM) and LCG (1 µM), titrated with H3 1-15 peptides either unmodified or bearing K9$(Me)_2$. The experiments were performed in 50 mM glycine, pH 9.15 at 25°C.

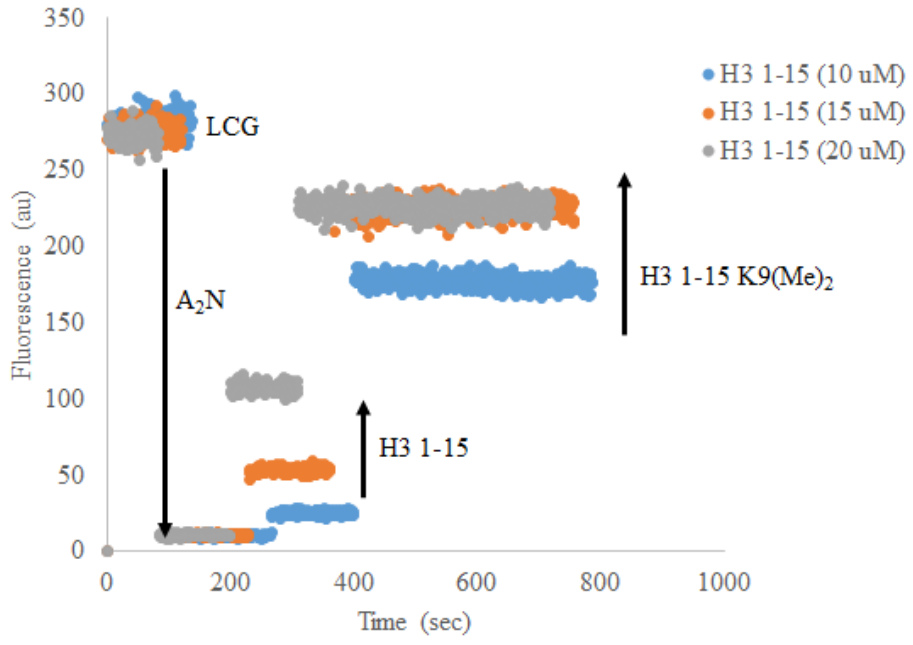


Figure 2.38 Cuvette based fluorescence differentiation assay. **A_2N** (15 µM) and LCG (1 µM), titrated with H3 1-15 peptides either unmodified or bearing K9$(Me)_2$. The experiments were performed in 50 mM glycine, pH 9.15 at 25°C.

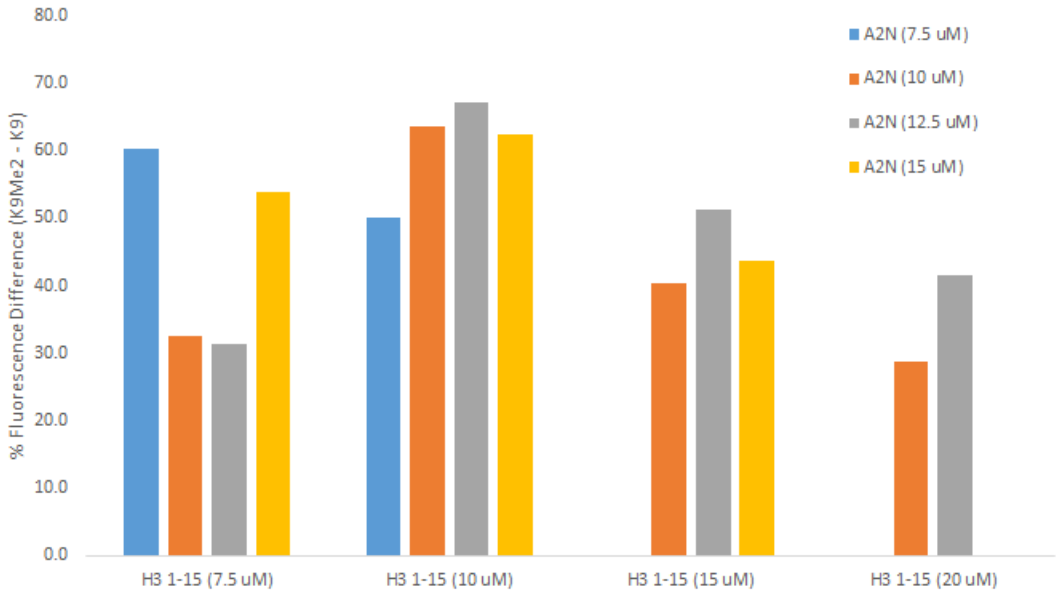


Figure 2.39 Combined fluorescence difference for **A_2N** (varying) and LCG (1 uM) sensor pair. Fluorescence difference calculated by subtracting the fluorescence of the unmethylated peptide (background) from the dimethylated peptide (signal) and dividing it by the initial LCG fluorescence to normalize across experiments.

2.5.5 Radioactive SAM assay

Reaction cocktails were prepared at a total volume of 10 µL in Eppendorf tubes. 7 µL of the reaction buffer was added, followed by 1 µL each of peptide and G9a at pre-determined concentrations for the specific assay. The reaction was initiated by pipetting ^{3}H-SAM (1µCi/uL) into the Eppendorf, followed by mixing by pipette aspiration. The Eppendorf tubes were placed in a water bath at 37°C and incubated for 45 minutes. Following reaction, the Eppendorf contents were transferred by pipette onto Whatman filter paper and washed 4 x 15 minute intervals in 50 mM Sodium Bicarbonate, with discarding of the wash between iterations. The filter papers were then dried on glass slides and transferred to scintiallation vials containing Ultima Gold F scintillation fluid. The vials were incubated for one hour and counted using a LKB Wallac 1214 Rackbeta liquid scintillation counter.

2.5.6 Cuvette based enzyme assay

Cuvette based fluorescent assays were conducted using a Cary Eclipse Fluorimeter using the Kinetic software suite. Experiments were run at 37° C using ex 370 em 510 with the PMT on high with a total cuvette volume of 100 μL. 80 μL of LCG was added to the cuvette and warmed to reaction temperature over two minutes. Fluorescence scanning was initiated and monitored for approximately 30 seconds prior to pausing and adding the receptor, with pipette aspiration to ensure complete mixing of the solution. Subsequent paused steps were done for the addition of peptide substrate, G9a enzyme, and finally S-Adenosylmethionine cofactor, after which the reaction was monitored for a total reaction time of fifteen to thirty minutes. Following reaction, the cuvette was rinsed 5x with DI water followed by 5x with methanol and was dried under nitrogen before running the next experiment.

REFERENCES

(1) Beaver, J.; Peacor, B.; Bain, J.; James, L.; Waters, M. *Org. Biomol. Chem.* **2015**, *13*, 3220.

(2) Ingerman, L. a; Cuellar, M. E.; Waters, M. L. *Chem. Commun.* **2010**, *46*, 1839.

(3) James, L. I.; Beaver, J. E.; Rice, N. W.; Waters, M. L. *J. Am. Chem. Soc.* **2013**, *135*, 6450.

(4) Pinkin, N. K.; Waters, M. L. *Org. Biomol. Chem.* **2014**, *12*, 7059.

(5) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838.

(6) Rougeulle, C.; Chaumeil, J.; Sarma, K.; Allis, C. D.; Reinberg, D.; Avner, P.; Heard, E. *Mol. Cell. Biol.* **2004**, *24*, 5475.

(7) Barski, A.; Cuddapah, S.; Cui, K.; Roh, T. Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. *Cell* **2007**, *129*, 823.

(8) Falkenberg, K. J.; Johnstone, R. W. *Nat. Rev. Drug Discov.* **2014**, *2*, 673.

(9) Timp, W.; Feinberg, A. P. *Nat. Rev.* **2013**, *13*, 497.

(10) Wang, G. G.; Allis, C. D.; Chi, P. *Trends Mol. Med.* **2007**, *13*, 363.

(11) Hake, S. B.; Xiao, a; Allis, C. D. *Br. J. Cancer* **2007**, *96*, R31.

(12) He, Y.; Korboukh, I.; Jin, J.; Huang, J. *Acta Biochim. Biophys. Sin.* **2012**, *44*, 70.

(13) Qian, C.; Zhou, M. M. *Cell. Mol. Life Sci.* **2006**, *63* , 2755.

(14) Dindar, G.; Anger, A. M.; Mehlhorn, C.; Hake, S. B.; Janzen, C. J. *Nat. Commun.* **2014**, *5*, 5313.

(15) Dirk, L. M. a; Flynn, E. M.; Dietzel, K.; Couture, J. F.; Trievel, R. C.; Houtz, R. L. *Biochemistry* **2007**, *46*, 3905.

(16) Wu, H.; Min, J.; Lunin, V. V.; Antoshenko, T.; Dombrovski, L.; Zeng, H.; Allali-Hassani, A.; Campagna-Slater, V.; Vedadi, M.; Arrowsmith, C. H.; Plotnikov, A. N.; Schapira, M. *PLoS One* **2010**, *5*, e8570

(17) Kwon, T.; Chang, J. H.; Kwak, E.; Lee, C. W.; Joachimiak, A.; Kim, Y. C.; Lee, J. W.; Cho, Y. *EMBO J.* **2003**, *22*, 292.

(18) Krishnan, S.; Horowitz, S.; Trievel, R. C. *ChemBioChem* **2011**, *12*, 254.

(19) Guo, H.-B.; Guo, H. *Proc. Natl. Acad. Sci.* **2007**, *104*, 8797.

(20) Zhang, X.; Bruice, T. C. *Proc. Natl. Acad. Sci.* **2008**, *105*, 5728.

(21) Bedford, M. T. *J. Cell Sci.* **2007**, *120*, 4243.

(22) Bedford, M. T.; Clarke, S. G. *Mol. Cell* **2009**, *33*, 1.

(23) Ramón-Maiques, S.; Kuo, A. J.; Carney, D.; Matthews, A. G. W.; Oettinger, M. a; Gozani, O.; Yang, W. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18993.

(24) Zhang, X.; Zhou, L.; Cheng, X. *EMBO J.* **2000**, *19*, 3509.

(25) Zhang, X.; Cheng, X. *Structure* **2003**, *11*, 509.

(26) Weiss, V. H.; McBride, a E.; Soriano, M. a; Filman, D. J.; Silver, P. a; Hogle, J. M. *Nat. Struct. Biol.* **2000**, *7*, 1165.

(27) Rust, H.; Zurita-Lopez, C.; Clarke, S.; Thompson, P. *Biochemistry* **2011**, *50*, 3332.

(28) Wang, M.; Xu, R.-M.; Thompson, P. R. *Biochemistry* **2013**, *52*, 5430.

(29) Wagner, T.; Jung, M. *Nat. Biotechnol.* **2012**, *30*, 622.

(30) Vedadi, M.; Barsyte-Lovejoy, D.; Liu, F.; Rival-Gervier, S.; Allali-Hassani, A.; Labrie, V.; Wigle, T. J.; Dimaggio, P. a; Wasney, G. a; Siarheyeva, A.; Dong, A.; Tempel, W.; Wang, S.-C.; Chen, X.; Chau, I.; Mangano, T. J.; Huang, X.-P.; Simpson, C. D.; Pattenden, S. G.; Norris, J. L.; Kireev, D. B.; Tripathy, A.; Edwards, A.; Roth, B. L.; Janzen, W. P.; Garcia, B. a; Petronis, A.; Ellis, J.; Brown, P. J.; Frye, S. V; Arrowsmith, C. H.; Jin, J. *Nat. Chem. Biol.* **2011**, *7* (8), 566.

(31) Kaniskan, H. Ü.; Szewczyk, M. M.; Yu, Z.; Eram, M. S.; Yang, X.; Schmidt, K.; Luo, X.; Dai, M.; He, F.; Zang, I.; Lin, Y.; Kennedy, S.; Li, F.; Dobrovetsky, E.; Dong, A.; Smil, D.; Min, S.-J.; Landon, M.; Lin-Jones, J.; Huang, X.-P.; Roth, B. L.; Schapira, M.; Atadja, P.; Barsyte-Lovejoy, D.; Arrowsmith, C. H.; Brown, P. J.; Zhao, K.; Jin, J.; Vedadi, M. *Angew. Chemie Int. Ed.* **2015**, *54*, 5166

(32) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.

(33) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41.

(34) Luo, M. *ACS Chem. Biol.* **2012**, *7*, 443.

(35) Suh-Lailam, B. B.; Hevel, J. M. *Anal. Biochem.* **2010**, *398*, 218.

(36) Furey, T. S. *Nat. Rev. Genet.* **2012**, *13*, 840.

(37) Gauthier, N.; Caron, M.; Pedro, L.; Arcand, M.; Blouin, J.; Labonte, a.; Normand, C.; Paquet, V.; Rodenbrock, a.; Roy, M.; Rouleau, N.; Beaudet, L.; Padros, J.; Rodriguez-Suarez, R. *J. Biomol. Screen.* **2012**, *17*, 49.

(38) Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D. *Curr. Biol.* **2011**, *21*, 53.

(39) Bradbury, A.; Plückthun, A. *Nature* **2015**, *518*, 27.

(40) Collazo, E.; Couture, J. F.; Bulfer, S.; Trievel, R. C. *Anal. Biochem.* **2005**, *342*, 86.

(41) Nguyen, B. T.; Anslyn, E. V. *Coord. Chem. Rev.* **2006**, *250*, 3118.

(42) Shaikh, M.; Mohanty, J.; Singh, P. K.; Nau, W. M.; Pal, H. *Photochem. Photobiol. Sci.* **2008**, *7*, 408.

(43) Nau, W. M.; Ghale, G.; Hennig, A.; Bakirci, H.; Bailey, D. M. *J. Am. Chem. Soc.* **2009**, *131*, 11558.

(44) Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. *ChemBioChem* **2010**, *11*, 63.

(45) Koh, K. N.; Araki, K.; Ikeda, a; Otsuka, H.; Shinkai, S. *J. Am. Chem. Soc.* **1996**, *118*, 755.

(46) Dsouza, R. N.; Pischel, U.; Nau, W. M. *Chem. Rev.* **2011**, *111*, 7941.

(47) Jin, T. *J. Incl. Phenom.* **2003**, *45*, 195.

(48) Doose, S.; Neuweiler, H.; Sauer, M. *ChemPhysChem* **2005**, *6*, 2277.

(49) Guo, D.-S.; Uzunova, V. D.; Su, X.; Liu, Y.; Nau, W. M. *Chem. Sci.* **2011**, *2*, 1722.

(50) Geddes, C. D. *Meas. Sci. Technol.* **2001**, *12*, R53.

(51) Patnaik, D.; Hang, G. C.; Estève, P. O.; Benner, J.; Jacobsen, S. E.; Pradhan, S. *J. Biol. Chem.* **2004**, *279*, 53248.

(52) Dillon, M. B. C.; Bachovchin, D. a.; Brown, S. J.; Finn, M. G.; Rosen, H.; Cravatt, B. F.; Mowen, K. a. *ACS Chem. Biol.* **2012**, *7*, 1198.

(53) Osborne, T. C.; Obianyo, O.; Zhang, X.; Cheng, X.; Thompson, P. R. *Biochemistry* **2007**, *46*, 13370.

(54) Butterfield, S. M.; Goodman, C. M.; Rotello, V. M.; Waters, M. L. *Angew. Chemie - Int. Ed.* **2004**, *43*, 724.

CHAPTER 3 Fingerprint Sensor Array for Combinatorial Histone Modifications

3.1 Introduction

Post-translational modifications of histone proteins form a complex landscape of markers, interacting with each other in a combinatorial language known as the 'histone code'.[1] However, this combinatorial nature can significantly complicate downstream screening and analysis. The two current leading methods for the end point detection of PTMs are antibodies and mass spectrometry (MS) proteomics, but the former suffers from issues pertaining to reproducibility and off target effects while the latter requires expensive equipment and sample preparation that not every lab has access to.[2,3] Because of this, there exists a need for facile high-throughput analysis method of combinatorial PTMs to allow further study of the complex interplay of the 'histone code'

We have previously described a series of small molecular receptors that are capable of recognizing methylated lysine and distinguishing between the varying degrees of methylation, as well as different degrees of arginine methylation.[4–7] Additionally, these receptors are capable of functioning in an indicator displacement assay with the small molecule fluorophore lucigenin (LCG). Of the three receptors studied, **A$_2$B**, **A$_2$D**, and **A$_2$N**, each produces a different signature of fluorescence recovery upon binding to methylated lysine. These fluorescence signals can be combined and analyzed using linear discriminate analysis (LDA) to produce a sensor array capable of fingerprinting each methylation state of lysine. Critically, each sensor exhibits different sensitivity to neighboring modifications on the histone peptide analytes, such that small perturbations in binding affinity are reflected in the fluorescence output. This allows the sensor

array to expand beyond lysine methylation and classify analytes bearing a larger subset of modifications, enabling the study of the complex 'histone code'.

3.2 Background

3.2.1 The Histone Code

Post-translational modifications are recognized by a large variety of different 'reader' proteins. However, studies have proposed that the marks may act in tandem to create a language or pattern of modification, dubbed the 'histone code' and shown in Figure 3.1.[1]

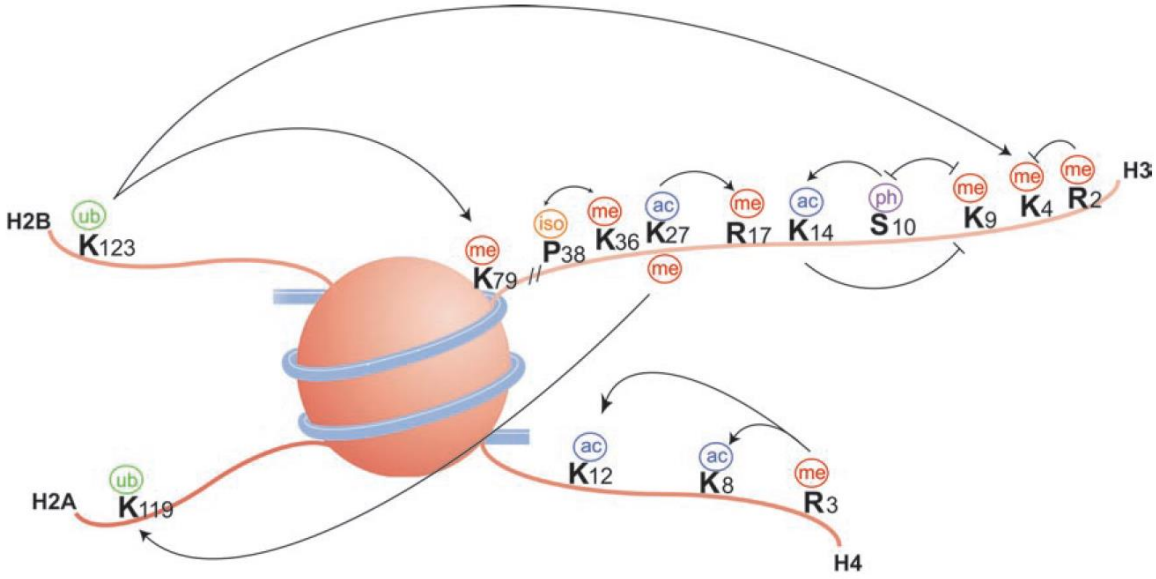


Figure 3.1 Examples of the 'histone code', the cross talk between modifications on the histone tail. Arrows indicate a positive cooperation while flat heads indicate an inhibitory effects.[8] Reprinted with permission from Macmillan Publishers Ltd: *Cell Research*, **2011**, *21*, 381, copyright 2011.

As shown above, PTMs can either cooperatively recruit complexes to further modify the histone tail, or can prevent enzymes and reader proteins from binding to the histone, facilitating different downstream events. In 2005, the Allis group showed that the recognition of lysine 9 trimethylation in the H3 tail by HP1 chromodomain, an important interaction for maintaining

heterochromatin, was perturbed by neighboring serine 10 phosphorylation, suggesting a regulatory mechanism.[9] Interestingly, the position of modification is just as critical as the identity. In the case of phosphorylation, recent work from the Chakravarti group has shown that the methylation of lysine 4 on H3 is proceeded by threonine 11 phosphorylation, which is responsible for recruiting the lysine methyltransferase.[10]

Multivalent binding regions in histone reader proteins have been observed recently as a major recognition motif.[11] For example, in 2007, the Yang group disclosed the structure of Recombination activating gene 2 (RAG2), which contains a PHD finger responsible for binding the trimethylated form of lysine 4 in the histone H3 tail.[12] However, they observed that the protein RAG2 was a mutation from the RAG1 domain, where an acidic side chain is substituted by a tyrosine, allowing recognition of and enhanced binding to an asymmetrically dimethylated arginine, where the RAG1 protein lacking this tyrosine is inhibited by neighboring methylation.

3.2.2 Tools for Studying the Histone Code

While understanding of the complexity of the epigenetic 'histone code' is rapidly increasing, there still exists the need for sensitive assays that are capable of rapidly and accurately classifying the landscape of modification. Of the available assays, antibody based high-throughput readouts and mass spectrometry are the current frontrunners.[13–16]

Antibody based methods are still the standard method of detecting and studying PTMs. Traditional methods such as ChIP-seq allow the direct analysis of which genes are modified by which PTMs.[13] Additionally, more high-throughput methods that utilize vast combinations of histone peptides have been used to reveal novel post-translational modifications.[17] However, recent studies have shown that commercial antibodies are potentially unreliable. The Strahl lab

has shown through peptide microarray analysis that commercial antibodies are susceptible to false results due to neighboring modifications, while the Lieb group has shown that in a panel of antibodies, 25% failed due to problems in specificity or utility.[2,3,18]

With multiple issues interfering with antibody effectiveness, proteomics methods have played a large role in the identification and study of PTMs. Mass spectrometry based methods provide an unbiased look at the modification landscape because they do not rely on binding to a specific epitope that can be disrupted. Additionally, the high sensitivity of modern MS instrumentation allows the unique ability to monitor the changes in a normal cell versus modified cellular states using quantitative methods such as SILAC, vastly increasing the knowledge of the PTM landscape.[15,19,20] However, the instrumentation required for such MS studies is still expensive and out of reach for many labs, and most PTMs are in such low abundance that they require some kind of amplification step to increase signal, returning to antibody or other enrichment methods that reduce the robustness of the technique.[21]

3.2.3 Sensor Arrays

Recently, the trend of sensing analytes has been transitioning from the traditional 'lock and key' method of designing the most selective single receptor to a more biomimetic approach, the generation and application of multiple general or less specific receptors.[22] This principle, known as the 'artificial nose' is advantageous because it combines a pattern of readout, allowing the study of analytes that were not designed as specific targets of the receptor, as shown in Figure 3.2.

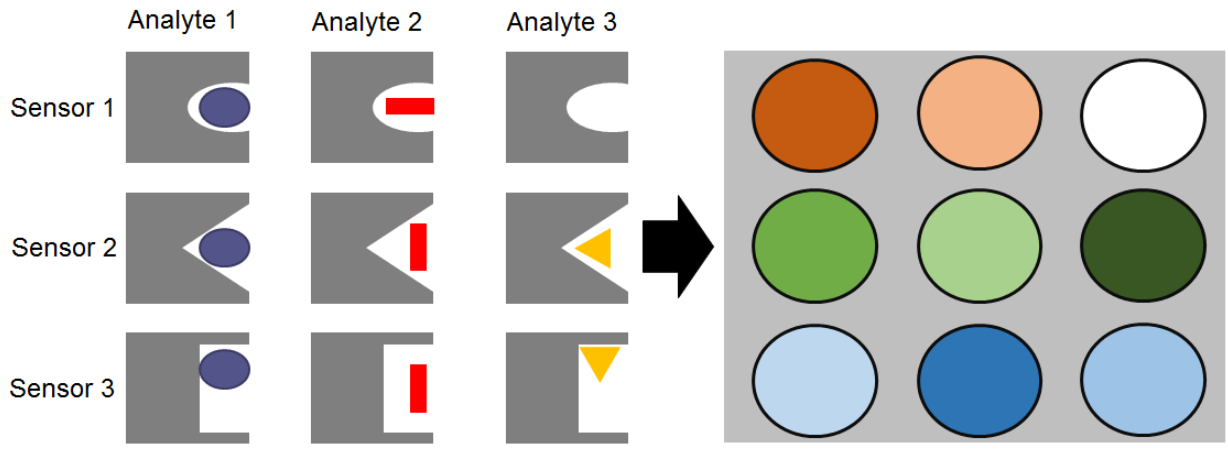


Figure 3.2 General assay design for pattern-based recognition and identification of analytes.

In this assay, receptors are designed to bind to one specific analyte, such as sensor 1 to analyte 1 or sensor 2 to analyte 3 (Figure 3.2). However, each receptor binds to each analyte with varying degrees of affinity and exhibits some response. Therefore, while classification may have been impossible with a single sensor, the pattern of signal allows for correct identification of each analyte.

This method of combinatorial sensor output has been used to great success by a number of groups. One of the more famous examples was conducted by the Anslyn group, in which they demonstrated the classification of red wine flavonoids using a sensor array of histidine bearing peptides, colorimetric indicators, and metal ions, as shown in Figure 3.3.[23]

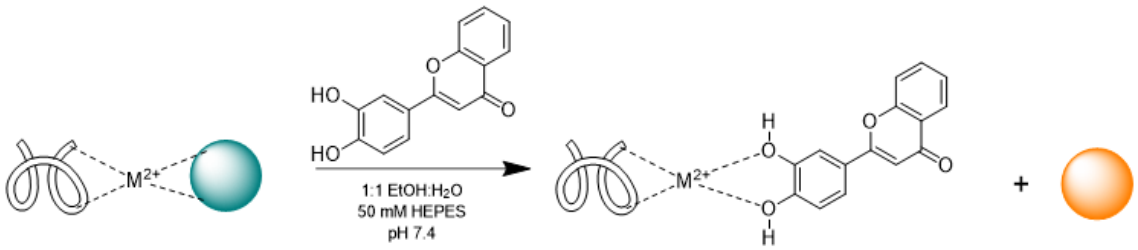


Figure 3.3 Sensor design for red-wine flavonoids.[23] Reproduced from *Chem Sci*, **2011**, *2*, 439 with permission of The Royal Society of Chemistry.

The group was able to utilize the different affinity of each flavonoid for the peptide complex to produce an indicator displacement assay, wherein the colorimetric signal could be calculated based on analyte binding. This allowed the classification of several different varietals of red wine based on the flavonoids present, a result much more difficult and costly to obtain if a specific receptor had to be designed for each separate analyte.

In 2014, the Anzenbacher group described the application of two fluorescent sensors based on cucurbituril motifs that were quenched by the binding of Eu^{3+}.[24] Upon binding of the cucurbiturils to basic amino acids, such as lysine and arginine, the Eu^{3+} was displaced, causing a recovery in fluorescence signal. The use of both sensors was key to the identification of ten different basic amino acid analytes, depending largely on the size complementarity of guest binding to hosts **1** and **2** (Figure 3.4). This enabled the classification of structurally similar analytes, again using an established sensor motif that is not designed to be highly selective.

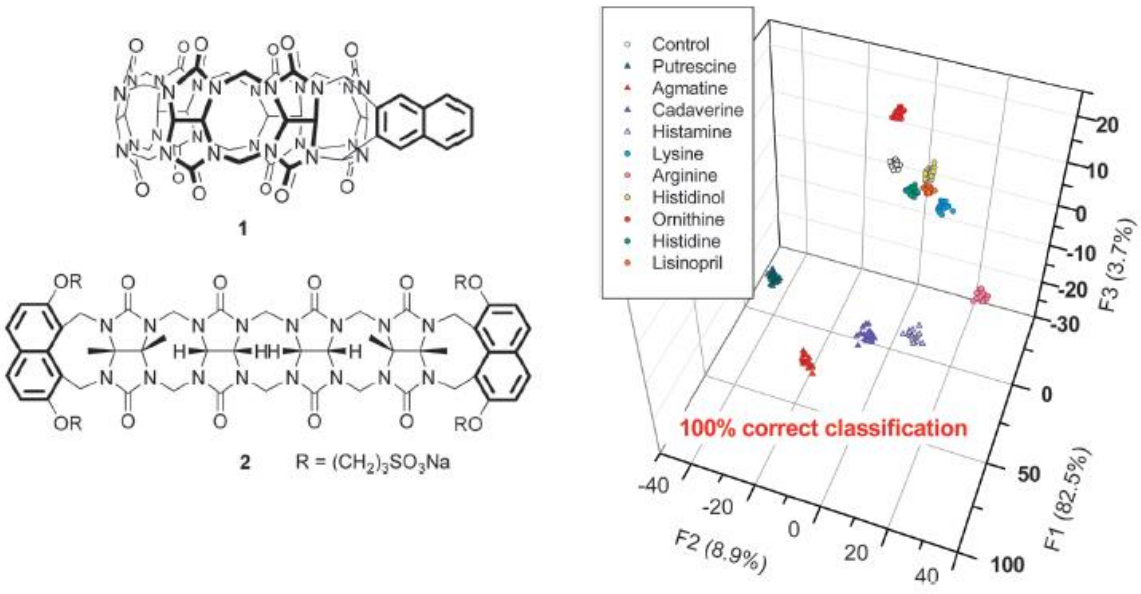


Figure 3.4 Structure of the fluorescent cucurbiturils **1** and **2** as well as the LDA score plot for the response of ten analytes to the sensors, demonstrating 100% accurate classification.[24]

Reproduced from *Chem. Commun*. **2014**, *50*, 61, with permission from The Royal Soceity of Chemistry.

As seen in Figure 3.4, and as is common for most combinatorial sensor arrays, the data output is often quite complicated. With multiple sensors giving multiple levels of response across multiple analytes, a traditional visual representation is not sufficient to convey the density of data. Therefore, statistical evaluation and interpretation of the data must take place prior to representation, and is most commonly performed using linear discriminate analysis (LDA).

LDA is a powerful statistical method able to explore the discriminatory capability of a given sensor array. It utilizes a jackknife or leave-one-out method to determine if a given array can correctly classify a group of analytes.[25] To generate the data, the statistical software converts the raw data, fluorescence in the case of an IDA, into a series of eigenvectors, which act as the canonical scores for each axis, allowing graphical representation.[26] A discriminant analysis (DA) puts the emphasis on clustering repetitive samples, such as multiple analyte trials. This is followed by separating the various clusters of samples, or classes, to achieve the best statistical relevance.[27]

In 2012, the Hof group utilized a calixarene (CX) based sensor array to study the classification of histone peptides bearing different modifications.[28] They utilized a series of three receptors, p-sulfanatocalix[4]arene (CX4), p-sulfanatocalix[6]arene (CX6), and the trisulfonated CX4 with a bromine substituted for a single sulfonate. These receptors, coupled with the fluorophore LCG, were able to distinguish between multiple analytes bearing single modifications such as acetylation or phosphorylation (Figure 3.5a). Additionally, they were able to distinguish the degrees of methylation (Figure 3.5b) as well as the site of trimethylation, suggesting that the differential sensing of the receptors is affected somewhat by the sequence neighboring the studied modification.

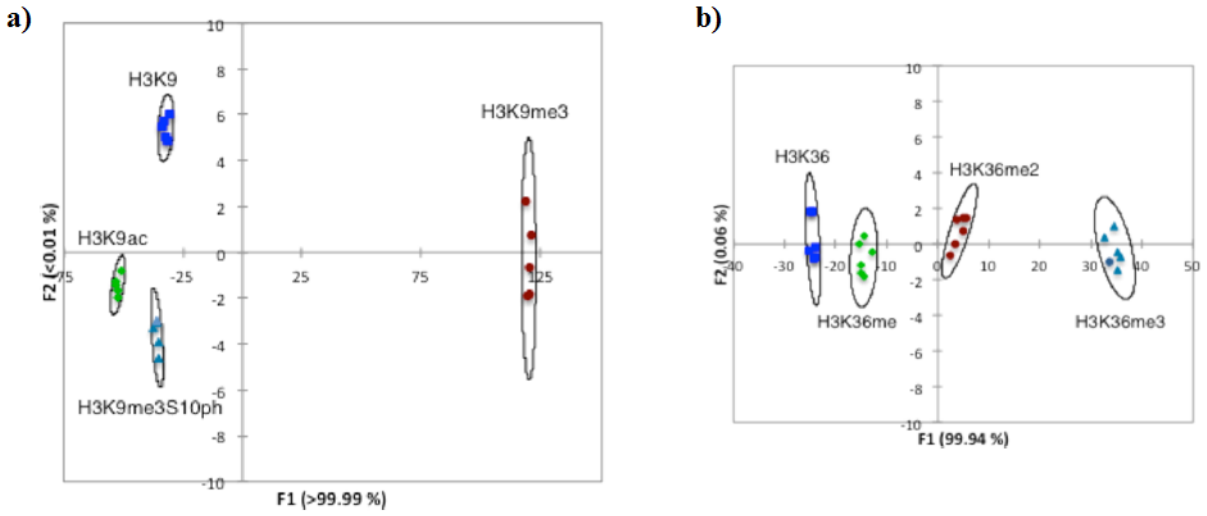


Figure 3.5 a) Calixarene sensor readout for histone peptides with a variety of modifications. b) Sensor readout for varying degrees of methylation on a histone H3 peptide.[28] Reprinted with permission from *J. Am. Chem. Soc.* **2012**, *134*, 11674. Copyright 2012 American Chemical Society.

In addition to single analyte sensing, the Hof group conducted experiments based on the theoretical sensing of enzymatic lysine methylation. The assay functioned on the principle of combining the readout of several concentration 'time points' of a mock enzyme run, such that it was possible to classify the percent conversion of an enzymatic reaction. However, the degree of differentiation using multiple receptors was at the limit of detection, as evidenced by the exceedingly small values on the F2 axis (Figure 3.5). Moreover, the ability to sense analytes bearing multiple modifications was still beyond the scope of the sensor array.

3.2.4 Motivation

The 'histone code' postulates that multiple modifications work in tandem to cause a downstream event, either by changing the physical properties of the chromatin structure, recruiting reader proteins that can associate with other modifiers, or signaling for 'writer' or 'eraser' proteins to further modify the neighboring sequence. However, as is evidenced in a

number of recent publications, the current tools for sensing this combinatorial landscape can be lacking, with significant cost associated to the analysis or problems associated with specificity.[18] Herein we envision a sensor array that can take advantage of binding to histone analytes in an indicator displacement assay system to produce a signal responsible for classifying the different modifications. Key to this work is the ability to detect and differentiate small perturbations in binding caused by the presence or absence of modifications adjacent to the recognized binding site, modulating the fluorescence response. Previous work has shown that the receptors have different binding affinities for $K(Me)_3$ based on the neighboring amino acid residues, and as such should be sensitive to the analytes described.[6,7] The ability to classify these types of peptides would expand the sensing ability to enzymatic reactions with substrates bearing multiple modifications, allowing an examination of the histone code hypothesis on actual enzymatic activity.

3.3 Results

3.3.1 Sensor Test With Short H3 Mimic Peptides.

We have previously established several receptors that bind to the fluorophore lucigenin and are able to differentially sense peptides bearing the methylation states of lysine. We observed that upon titration with identical peptides each receptor/dye pair had a different response, as seen in Figure 3.6.

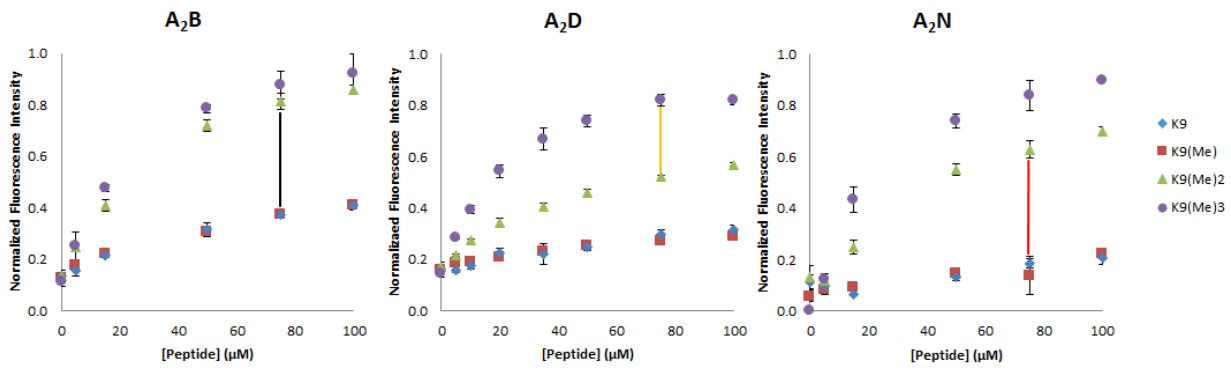


Figure 3.6 Fluorescence displacement for three receptor/LCG pairs by the histone H3 peptide Ac-YGGG-QTARK(Me)$_n$STG-NH$_2$. The colored bars on each titration signify the distinguishing fluorescence difference for each sensor.

Even within the relatively simple peptide sequence used in these titrations, each receptor had a different pattern of response. A_2B was best at identifying the higher order methylations, though was not sensitive to K(Me)$_2$ over K(Me)$_3$. A_2D was good at differentiating those two modifications, and A_2N displayed the best signal between K(Me)$_2$ and the lower methylation, while still maintaining some discrimination between K(Me)$_2$ and K(Me)$_3$.

The sensor array design incorporates the fluorescence signal from all three receptors in order to differentiate between peptide analytes. In a 96 well plate format, we can study multiple replicates of the same peptide at a fixed concentration in order to determine the sensitivity and classification accuracy of the system in a high-throughput manner. The initial experiments tested the discriminatory ability of all three sensors in this plate format using the short H3 peptides described previously.

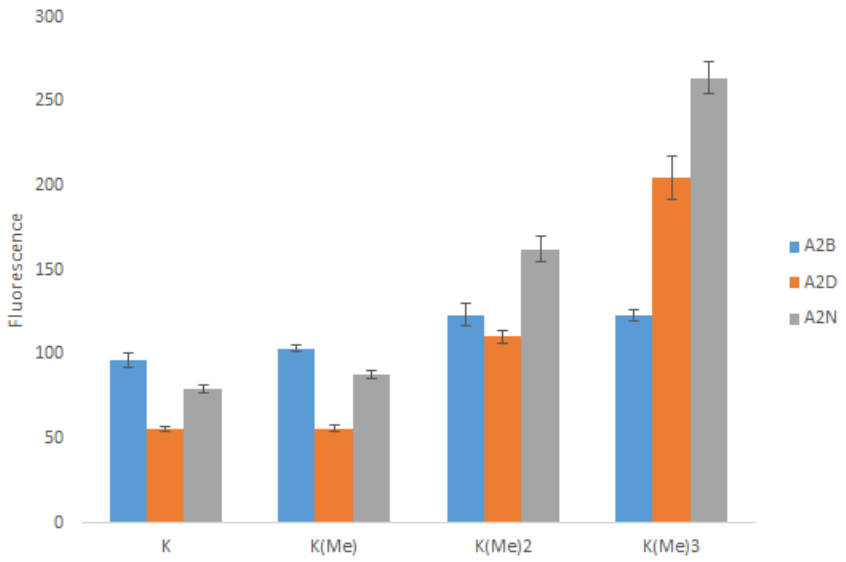


Figure 3.7 Fluorescence response of three sensors (**A_2B/LCG**, **A_2D/LCG**, **A_2N/LCG**) to the short H3 peptide Ac-YGGG-QTARK(Me)$_n$STG-NH_2 (30 μM) in 50 mM glycine buffer, pH 9.15.

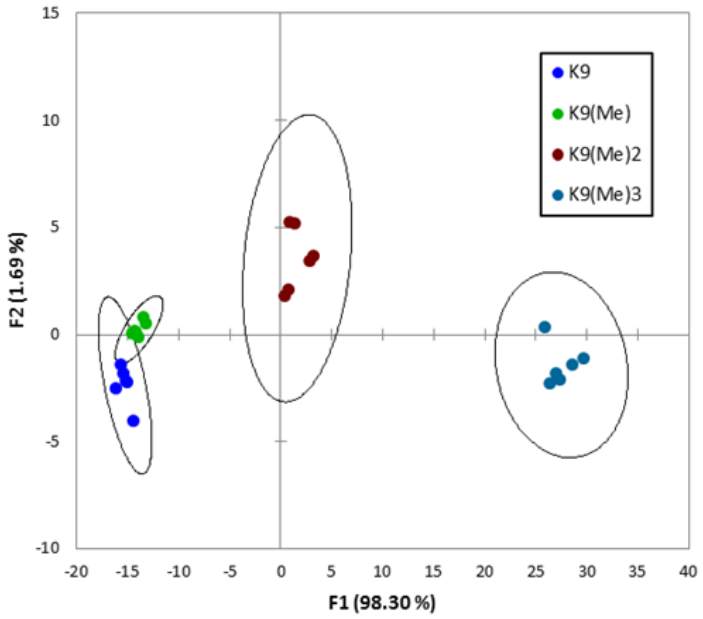


Figure 3.8 LDA plot of the fluorescence response to the short H3 peptides Ac-YGGG-QTARK(Me)$_n$STG-NH_2 (30 μM). Confidence ellipses at 95%.

The LDA plot of the four peptides shows clear classification into the corresponding methylation states of lysine, though there is some overlap in the 2D confidence ellipses at 95%.

However, each factor has a percentage next to it, representing how much of the classification power is coming from that specific eigenvalue score. This signifies that the x-axis eigenvalue, F1, is dominating the classification power of the assay, which is clear from the wide separation on this axis. Indeed, the F2 axis is only responsible for separating K and K(Me). This result demonstrates that the full power of the sensor array is not being utilized, three receptors are not needed to classify these short peptides. The raw fluorescence data agrees with this, in Figure 3.8 a quick visual examination of the graphs can distinguish each peptide, again with only K(Me) and K being difficult to identify.

To expand on the sensor array platform and take advantage of its discriminatory capacity, we expanded the analyte selection to the first 1-12 amino acids of the H3 tail, the unmodified version of which is Ac-ARTKQTARKSTGY-NH_2, adding tyrosine for accurate concentration determination. This sequence is unique because it has several possible sites of modification, two lysines, two arginines, two threonines, and one serine, all of which can host PTMs. To test this sequence, we synthesized four peptides bearing competitive methylation marks that our receptors should bind very tightly to. These peptides, shown in Table 3.1, are representative of cross-talk between methylation of lysine and arginine, and are a more difficult class of analyte to study.

Table 3.1 Peptides used in LDA for combinatorial methylation study

Abbreviation	Peptide Sequence
R2(Me)2a	Ac-A**R(Me)$_{2a}$**TKQTARKSTGY-NH2
K9(Me)3	Ac-ARTKQTAR**K(Me)$_3$**STGY-NH2
R2(Me)2a K9(Me)2	Ac-A**R(Me)$_{2a}$**TKQTAR**K(Me)$_2$**STGY-NH2
R2(Me)2a K9(Me)3	Ac-A**R(Me)$_{2a}$**TKQTAR**K(Me)$_3$**STGY-NH2

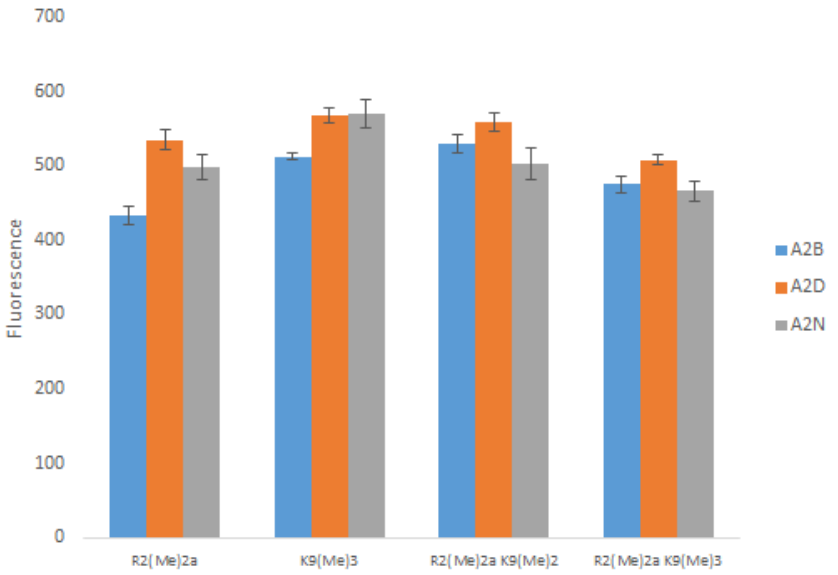


Figure 3.9 Fluorescence response of three sensors (**A_2B/LCG**, **A_2D/LCG**, **A_2N/LCG**) to the peptides in **Table 3.1** (30 μM) in 50 mM glycine buffer, pH 9.15.

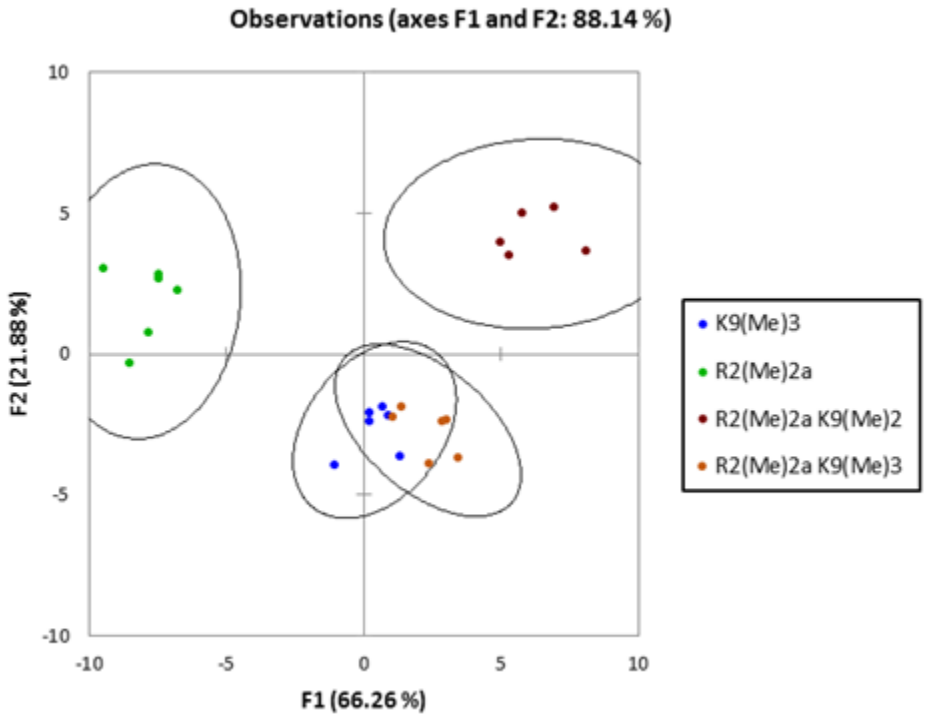


Figure 3.10 LDA plot of the fluorescence response to the peptides in **Table 3.1** (30 μM). Confidence ellipses at 95%.

The fluorescence plot in Figure 3.9 is drastically different from that of the short peptides, as would be expected for this class of analyte. We observed that the peptides were almost completely displacing the fluorophore, suggesting tight binding affinity and probably saturation of the receptor system, because we were operating at significantly high concentration. In a quick visual examination of the fluorescence output, it is difficult to distinguish the analytes, however LDA analysis again saw good classification of the species, though there is overlap amongst the two $K(Me)_3$ species. Gratifyingly we observed that we were beginning to harness more of our receptor discriminatory ability, with F1 loading at 66% and F2 loading at 22%. Interestingly, these two factors only comprise 88% of the data, with the remaining 12% located in factor F3. This extra factor provides the capacity to fully distinguish the analytes in three dimensions, as shown in Figure 3.11.

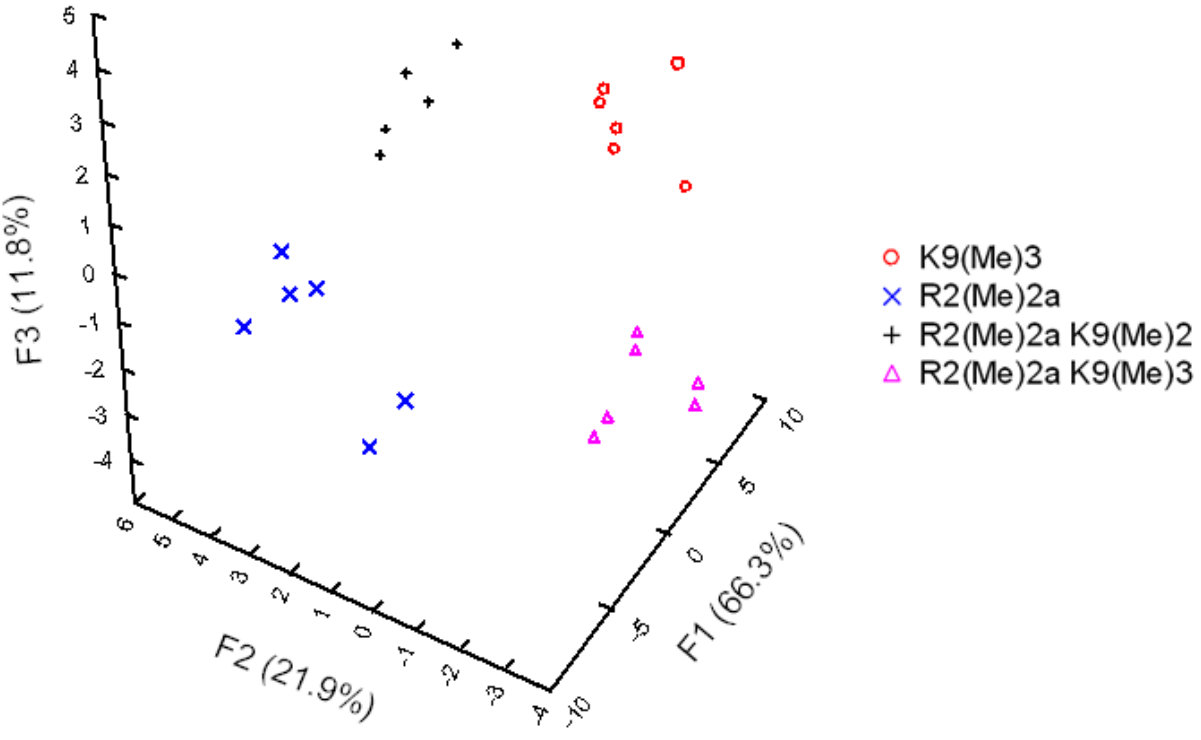


Figure 3.11 3D LDA plot of the fluorescence response to the peptides in **Table 3.1** (30 μM).

The three dimensional plot of the four peptides shows clear distinction between each sample, with the apparent overlap between the $K(Me)_3$ peptides gone due to the z-axis separation of the peptides. This result demonstrates the utility of the three receptor sensor array in distinguishing between complex histone peptides bearing multiple modifications, and suggests that the complexity can still be increased.

3.3.2 Expansion of Sensors

With the goal of studying even more complex peptides, we sought to expand the number of sensors available to us. We had previously described the fluorophore acridine orange (AO) (see Chapter 2) binding to **A_2B**. While this fluorophore was photobleached over the course of a real-time experiment, it should excel in an end point, single excitation assay. To test this, we examined whether the other receptors could successfully quench AO in the glycine buffer system.

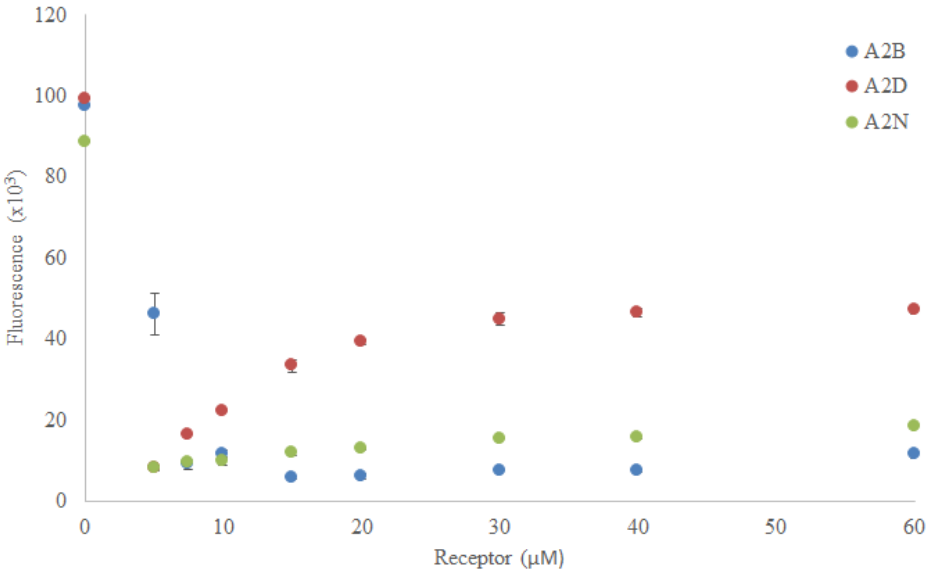


Figure 3.12 Fluorescence quenching of AO (5 μM) by three receptors in 50 mM glycine buffer, pH 9.15.

As seen in Figure 3.12, each receptor quenched AO fluorescence, with A_2D recovering fluorescence after the initial quenching through a mechanism hypothesized previously where two equivalents of **A_2D** can bind to one molecule of AO, creating a more hydrophobic environment and recovering fluorescence. Interestingly, **A_2B** did not experience this recovery contrary to previous results in a 10 mM Na_2HPO_4 buffer, pH 8.5, suggesting that buffer medium can influence the association. Each receptor did experience an initial fluorescence quenching, signifying that at fixed concentrations of AO and receptor they could act as useful sensors.

While the addition of AO should add more discriminatory factors to the array, it might not contribute much to the classification, because the dye/receptor K_d is similar to the LCG/receptor, suggesting that there won't be much difference in the signal response between the two systems. Therefore, we also wanted to add an additional receptor to the array. We had previously described the sensor **A_2G**, which was able to differentiate $K(Me)_3$ from $K(Me)_2$ in addition to providing high selectivity over lysine and minimal interference from neighboring residues.[7] We screened **A_2G** with lucigenin and found that it did bind and quench the fluorophore similarly to the other receptors, with a K_d of around 8 μM.

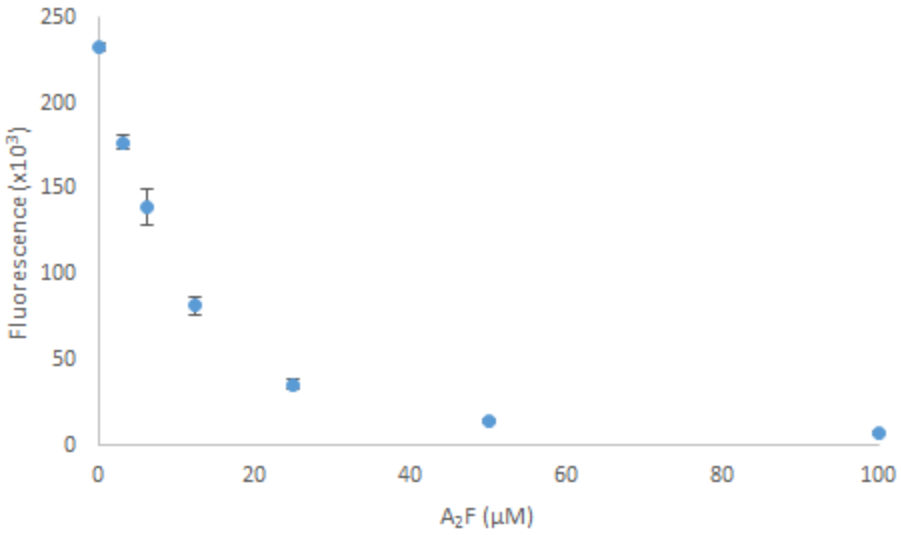


Figure 3.13 Fluorescence quenching of LCG (1 μM) by **A_2G** in 50 mM glycine buffer, pH 9.15.

We examined the inclusion of these new sensors into the discrimination of peptides bearing multiple methylation marks at multiple sites, which can be seen in Table 3.2.

Table 3.2 H3 peptides methylated to varying degrees at varying sties.

Abbreviation	Peptide Sequence
R2(Me)2a K9(Me)2	Ac-AR(Me)2aTKQTARK(Me)2STGY-NH2
R2(Me)2a K9(Me)3	Ac-AR(Me)2aTKQTARK(Me)3STGY-NH2
R2(Me)2a K4(Me)2	Ac-AR(Me)2aTK(Me)2QTARKSTGY-NH2
R2(Me)2a K4(Me)3	Ac-AR(Me)2aTK(Me)3QTARKSTGY-NH2

Our first examination was of the AO sensors in isolation. Using a 384 well plate, we could rapidly examine peptide fluorescence with minimal material cost, allowing the quick screening of conditions.

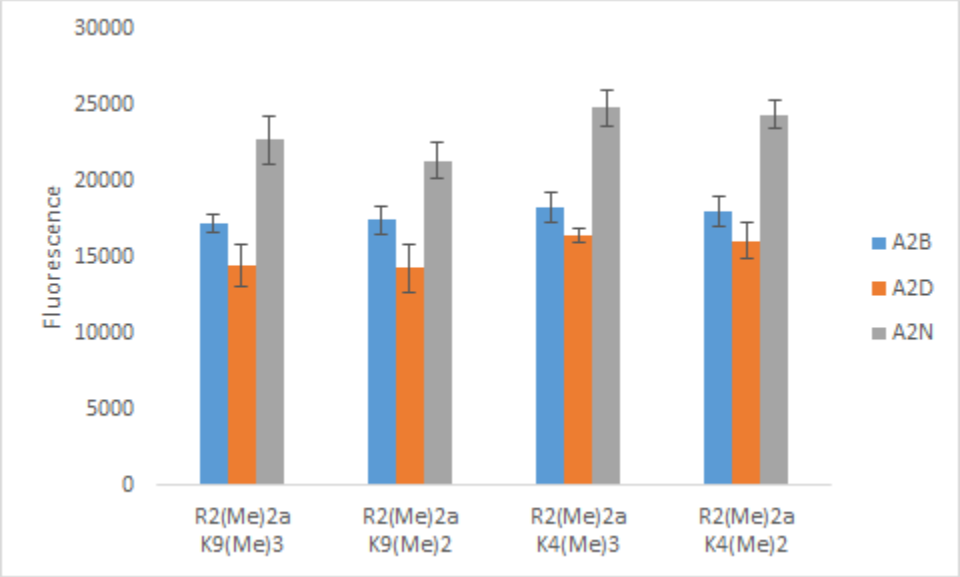


Figure 3.14 Fluorescence titration of AO sensors with the multiply methylated peptides in **Table 3.2** (15 μM) in 50 mM glycine buffer.

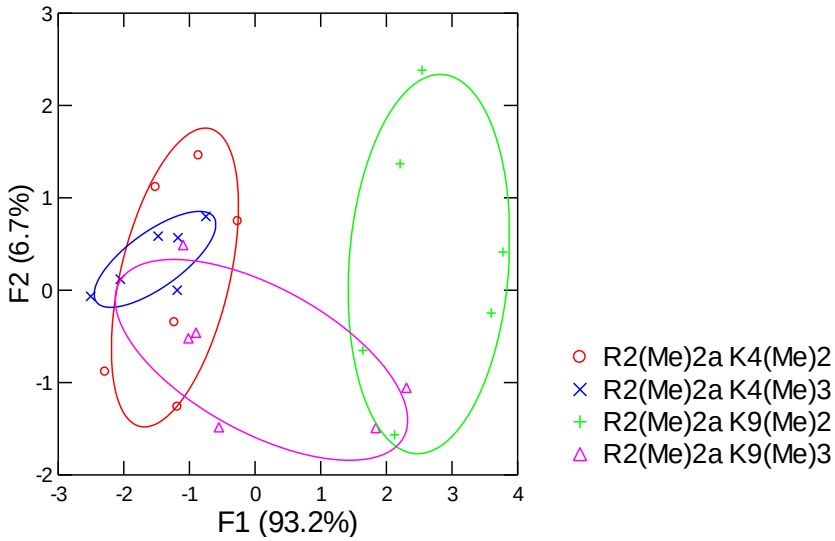


Figure 3.15 LDA of the fluorescence titration of AO sensors with the multiply methylated peptides in **Table 3.2** (15 μM) in 50 mM glycine buffer.

Upon analysis of the two figures above, it is clear that the AO sensor system is incapable of distinguishing between the four analytes described. It is not sensitive enough to the subtle changes in peptide sequence and methylation to allow for classification, with the LDA plot giving only 63% accuracy initially and 25% accuracy in the jackknife analysis. Therefore we did

not further purse the AO sensor system with methylated peptides, though it could still have utility for other classes of analytes.

3.3.3 Sensor Array of Single PTMs

With four sensors available, comprising of **A_2B**, **A_2D**, **A_2N**, and **A_2G**, we sought to expand the analyte classification capability of the array. We began by examining peptides bearing single modifications, though expanding those modifications away from the traditional methylation states studied thus far, as shown in Table 3.3.

Table 3.3 H3 peptides with single PTMs.

Abbreviation	Peptide Sequence
H3 1-12	Ac-ARTKQTARKSTGY-NH2
K9(Me)3	Ac-ARTKQTARK(Me)3STGY-NH2
K4Ac	Ac-ARTK(Ac)QTARKSTGY-NH2
R2(Me)2a	Ac-AR(Me)2aTKQTARKSTGY-NH2
T11 Phos	Ac-ARTKQTARKST(Phos)GY-NH2

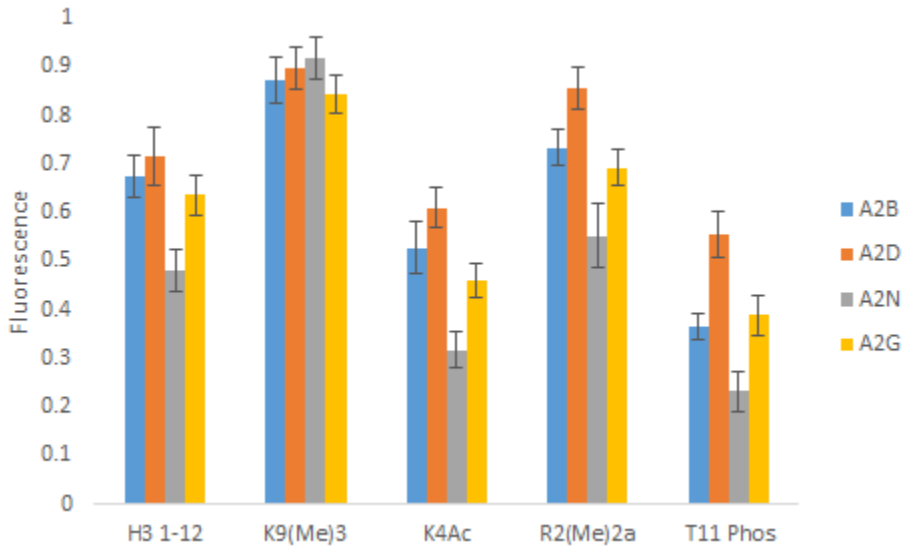


Figure 3.16 Fluorescence titration of LCG sensors with the single modification peptides in **Table 3.3** (15 μM) in 50 mM glycine buffer.

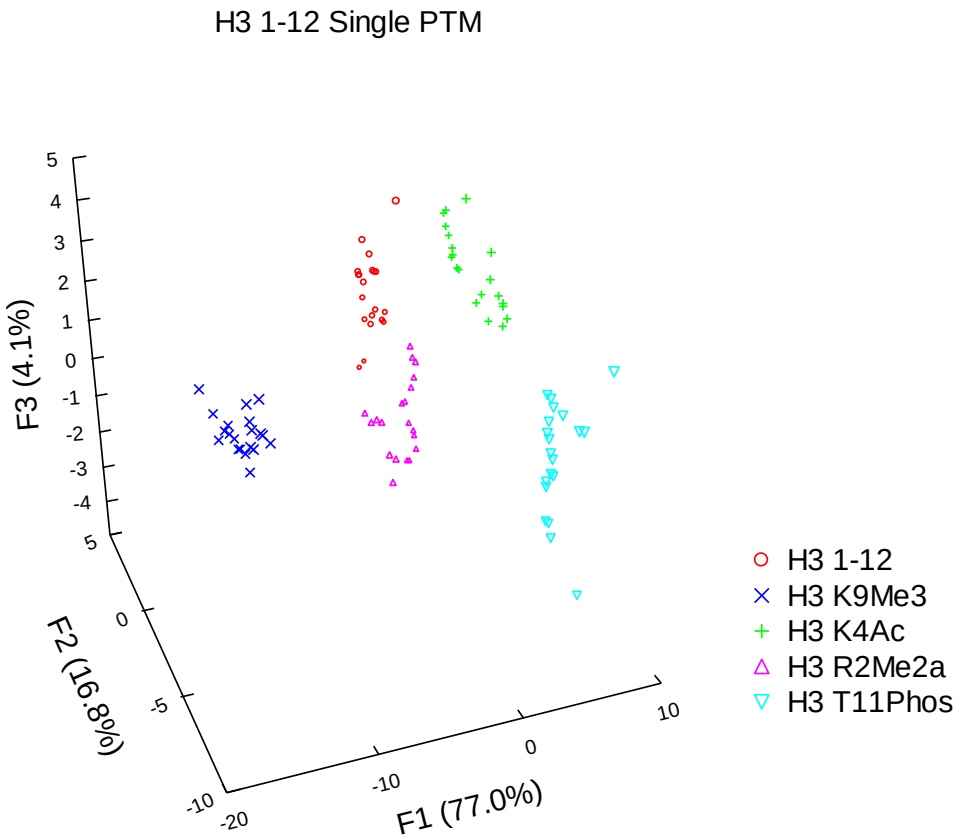


Figure 3.17 LDA of the fluorescence titration of LCG sensors with the single modification peptides in **Table 3.3** (15 μM) in 50 mM glycine buffer.

As evidenced above, the sensor array was successful at distinguishing between the peptides bearing single modifications, with 100% classification accuracy and 100% in the jackknife test. In this instance, plotting the F3 axis was necessary to prevent overlap of species. The indicator displacement can also be viewed as representative of analyte binding, exploring the effect modifications have on binding. Though the sensor array doesn't give thermodynamic data such as binding affinity, it does show general trends. For instance, the **A_2D** receptor, originally designed to bind asymmetric dimethyl arginine gave the highest signal for that peptide, though its signal was identical to the $K(Me)_3$ peptide, analogous to the binding affinity previously observed.[5] Interestingly, the sensors were sensitive to modification expected to hamper binding, the acetylation and phosphorylation, and though the pattern of response was similar for each of

these species (Figure 3.16), the phosphorylated peptide interacted less favorably with the receptors, suggesting that it is a more interfering modification for host binding.

3.3.4 Sensor Array for Peptides with Multiple Methylation

With all four sensors operational, we re-examined the ability to differentiate between peptides bearing not just multiple methylations, but also methylations in different positions on the peptide chain. Therefore, we examined four peptides, all with arginine asymmetric dimethylation concurrent with lysine methylation on either K4 or K9, as seen in Table 3.2.

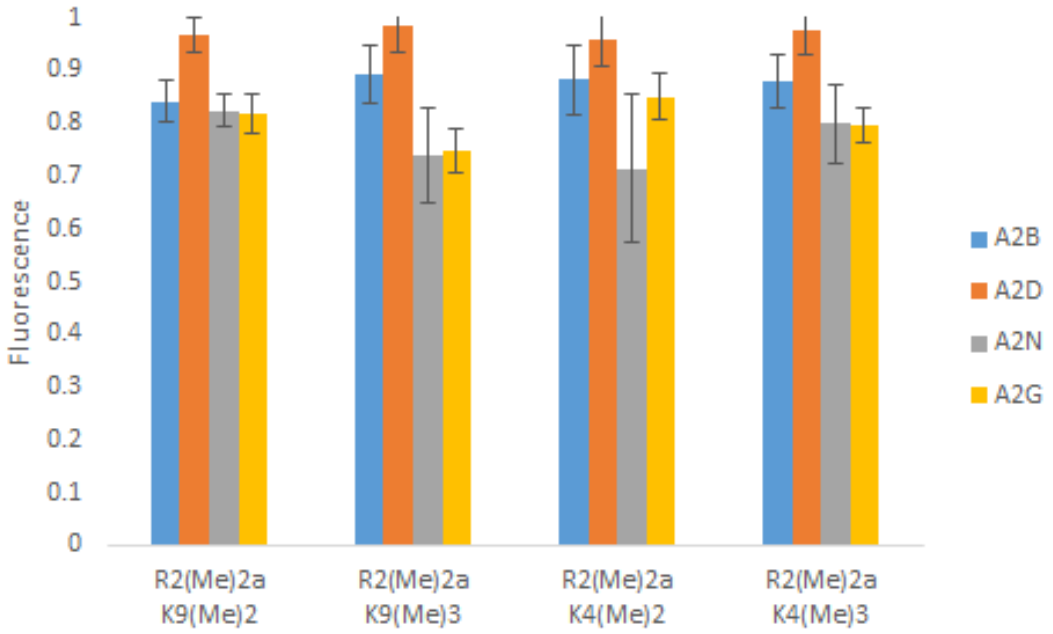


Figure 3.18 Fluorescence titration of LCG sensors with the multiply methylated peptides in **Table 3.2** (15 μM) in 50 mM glycine buffer.

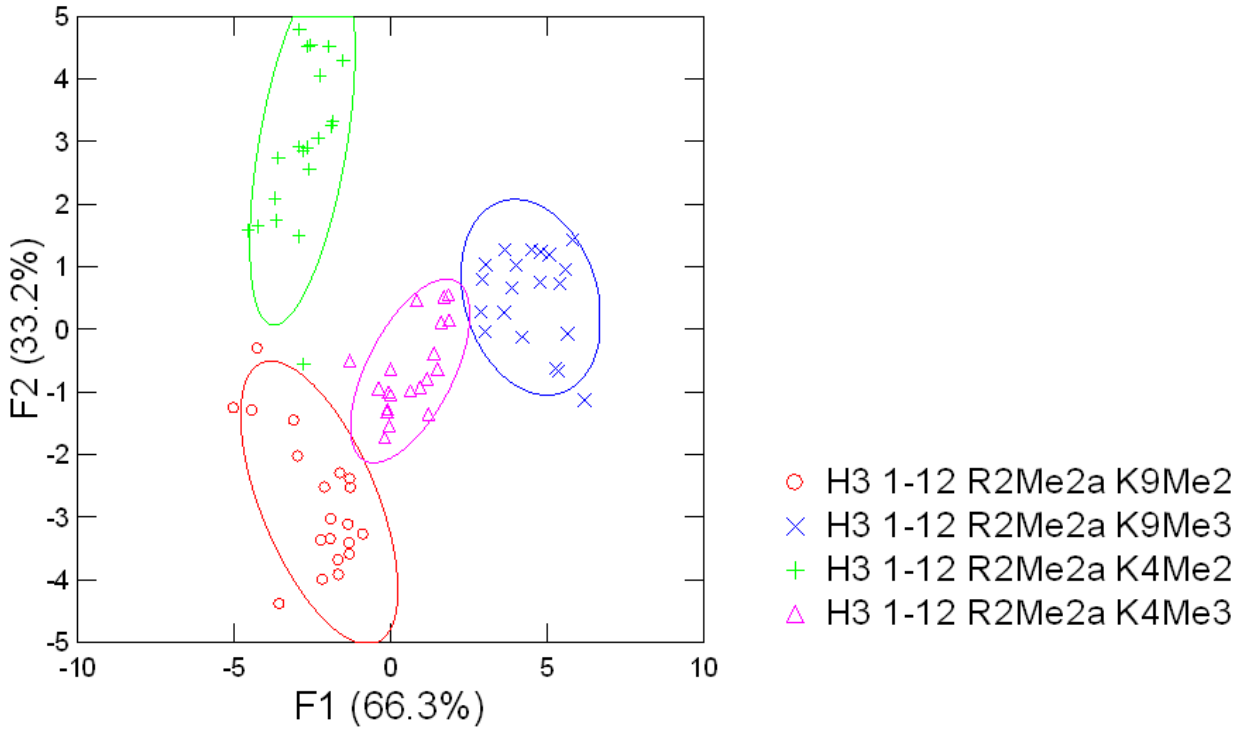


Figure 3.19 LDA of the fluorescence titration of LCG sensors with the multiply methylated peptides in **Table 3.2** (15 μM) in 50 mM glycine buffer. 85% confidence ellipses.

The sensor array was able to classify the multiply methylated peptides with 99% accuracy, with a 98% jackknife. As seen in Figure 3.19, the 99% was due to a single replicate of the R2(Me)$_{2a}$K4(Me)$_2$ being misclassified as R2(Me)$_{2a}$K9(Me)$_2$, meaning that it was purely a sequence misclass and not methylation state, a good result. The result of this fluorescence assay was quite surprising, not only can the sensor distinguish between K(Me)$_2$ and K(Me)$_3$ in the presence of a competitive arginine methylation, but it can classify where this modification is on the histone H3 tail. The similar modifications were grouped together in the assay, suggesting that the fluorescence response was corresponding to the binding affinities observed for lysine methylation in isolation.

3.3.5 Sensor Array for Peptides with Neighboring Phosphorylation

The next set of peptides we wanted to examine bore phosphorylation adjacent to lysine methylation. Phosphorylation is commonly known as an inhibitor of lysine methylation when adjacent, most likely due to the addition of negative charge interfering with peptide docking to the methyltransferase.[8] We hypothesized that this inhibition mechanism would be function within the sensor array as well, due to the negative charge of the receptors. However, because the sensor array does not require tight affinities, as long as the receptors still interact differentially, so phosphorylation does not abolish the previously observed selectivity, then the signal recovery should still be enough to classify the peptides. We studied four different peptides, each phosphorylated at threonine 11, with methylation either two residues away at lysine 9, which should have a highly repressive effect on binding, or 7 residues away at lysine 4, which should minimize the inhibition.

Table 3.4 H3 peptides with threonine phosphorylation and lysine methylation.

Abbreviation	Peptide Sequence
K9(Me)2 T11Phos	Ac-ARTKQTARK(Me)2ST(Phos)GY-NH2
K9(Me)3 T11Phos	Ac-ARTKQTARK(Me)3ST(Phos)GY-NH2
K4(Me)2 T11Phos	Ac-ARTK(Me)2QTARKST(Phos)GY-NH2
K4(Me)3 T11Phos	Ac-ARTK(Me)3QTARKST(Phos)GY-NH2

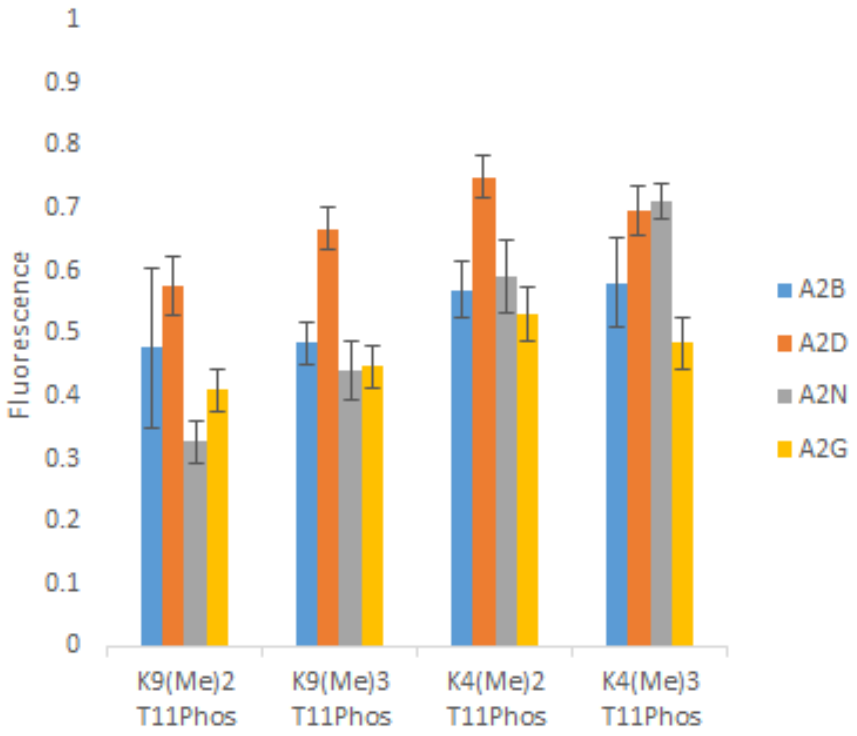


Figure 3.20 Fluorescence titration of LCG sensors with the phosphorylated peptides in **Table 3.2** (15 μM) in 50 mM glycine buffer.

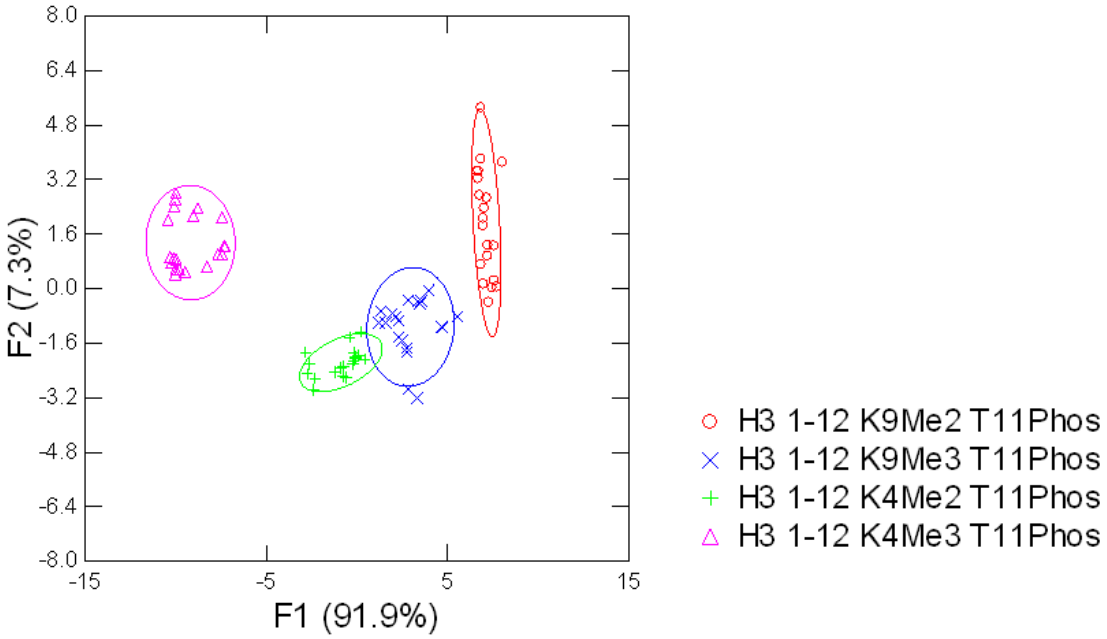


Figure 3.21 LDA of the fluorescence titration of LCG sensors with the phosphorylated peptides in **Table 3.2** (15 μM) in 50 mM glycine buffer. 85% confidence ellipses.

Once again, the sensor array was able to 100% classify the four peptides, with 100% in the jackknife as well. Also as expected, the phosphorylation had a much larger impact on the binding of the lysine 9 peptides, where it was only two residues away, suggesting that the anionic modification weakened the affinity of the methyl lysine binding through an electrostatic interaction. However, even with this effect, the binding selectivity still remained, with the $K9(Me)_3$ peptide binding slightly tighter to the receptors, causing an increased fluorescence signal. For the K4 position, the phosphorylation still weakened binding, though not to the same degree, suggesting a more long range interaction with the receptors and the H3 1-12 peptide.

3.3.6 Combined Sensor Array Output

Having demonstrated the utility of the sensor array for several classes of peptides, we wanted to examine its utility in peptide classification across functionality. This involved analysis of each previously run peptide compared to each of the others, for a total analyte sample set of 13 peptides at 20 replicates each.

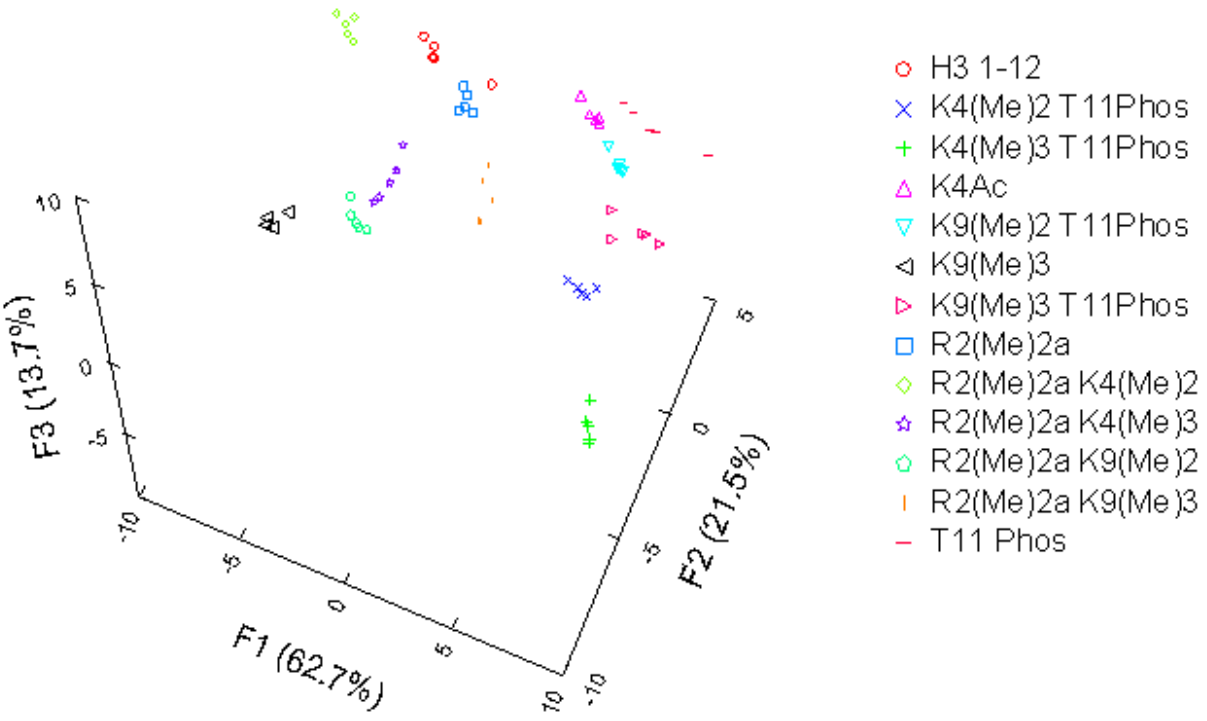


Figure 3.22 LDA of the combined sensor array output for all thirteen peptides. The first five data points for each analyte are displayed to lower the complexity of the output.

With thirteen different peptides the resulting graphical output is understandably complex. Figure 3.22 shows the first five replicates out of twenty in order to declutter the output. However, with all twenty replicates the assay is able to correctly classify the analytes with 96% accuracy, with 95% accuracy in the jackknife. Importantly, that percentage does not come from a large misclassification of a single analyte, each cluster gives a classification of at least 90%, showing that only one or two of the replicates are being mistakenly identified. This classification of a mass group of analytes is remarkable, especially since the receptors were designed to bind only to methylated lysine. In order to fully evaluate the robustness of the assay, a training/test set analysis was conducted using 50% of the replicates as the training set. This training set population was randomized each time to ensure full coverage of training/test set members. In this evaluation the training set correctly classified 96 ± 1% of the test set. This demonstrates a robust

analytical result, showing that the sensor array is able to analyze a large number of peptides with high accuracy.

The results highlight how a small molecule sensor array has advantages over the traditional methods of analyzing combinatorial histone modifications. Proteomics mass spectrometry is capable of classifying peptides, but the time commitment and expense of instrumentation is intense. Here we can analyze each analyte with multiple replicates using minimal material in only 30 minutes per experiment, using a simple fluorescence plate reader. Additionally, the sensor array is more suited to multiply modified peptides compared to the analogous antibodies. Whereas the high specificity of antibodies could prevent binding to the target epitope in the presence of a nearby modification, the receptor platform is simply perturbed by this interaction, utilizing the new affinity to correctly classify each species.

3.4 Application of the Sensor Array to Enzymatic Reactions

Having shown the utility of a small-molecule receptor based sensor arrays for end-point analysis of histone peptides we wanted to expand the platform to enzymatic reactions. The advantage of this approach for studying enzymes is that the researcher is not limited to the designed targets of the sensor, meaning that the assay can expand beyond histone methylation as a target, as long as the substrate and product of the reaction will have differential binding to the variety of receptor/dye pairs. This enables the study of more complex histone peptides as well as unique reactions. We used a histone kinase as the example enzymatic reaction studied in this experiment. Kinase assays typically rely on quantification of radioactive phosphorus incorporation, but this is a dangerous and expensive method, as ^{32}P has a half-life of only 15 days, requiring consistent stock replenishment.

Our proof of concept for enzymatic assays involved a simple plate based experiment that functions as a mimic for an actual enzyme. For the actual enzyme assay, the sensor platform needs to be 'trained', wherein concentration based 'time-points' are analyzed to create a response pattern assuming substrate depletion with concurrent product formation. For our assay this manifested in four distinct experiments, one plate at 100% substrate, representing the state prior to enzymatic conversion, one plate at 100% of product formation to define the reaction end point, as well as two plates for the intermediate reaction points, 33% and 66% product formation respectively.

The first mock enzyme experiment was based on the kinase PKN1, which is responsible for phosphorylation of threonine 11 in the histone tail.[10] To simulate this enzyme's activity, we used the H3 1-12 peptide (Ac-ARTKQTARKSTGY-NH_2) and the H3 1-12 T11 Phos peptide (Ac-ARTKQTARKST(Phos)GY-NH_2).

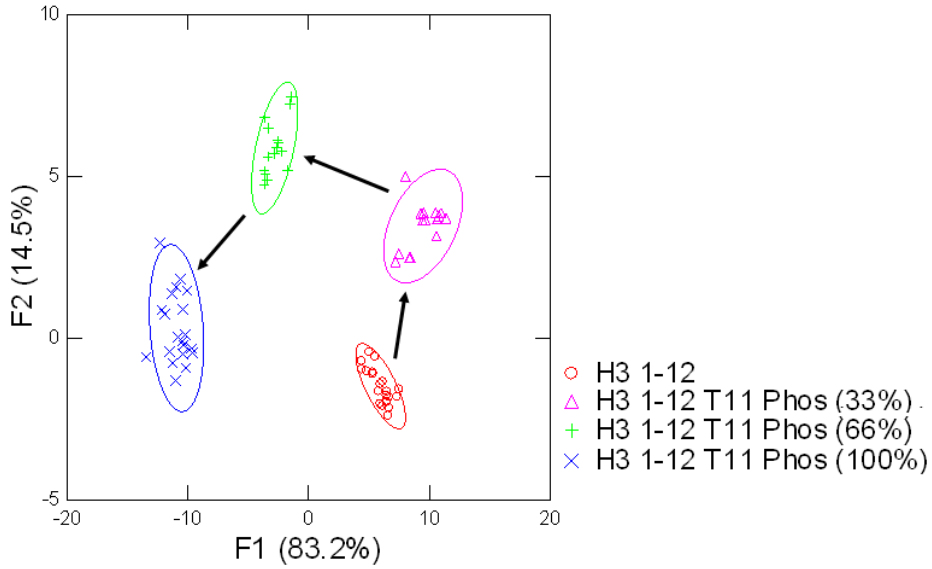


Figure 3.23 LDA of the mock enzymatic kinase reaction monitoring conversion of H3 1-12 to H3 1-12 T11 Phos. The substrate and product were both at 15 µM, with 33% conversion at 5:10 µM substrate:product and 66% at 10:5 µM. Arrows were added to represent the path of phosphorylation, confidence ellipses at 90%.

Figure 3.23 shows the result of the 'training set' of the phosphorylation of the H3 1-12 peptide. The assay is able to discriminate each concentration point with 100% accuracy and 100% jackknife compatibility. This training set could be applied to a real enzyme by running the kinase alongside the training set, then inputting the resulting fluorescence for each sensor as a blind data point. This data point should associate with one of the regions of enzymatic conversion, allowing monitoring of the reaction progress. Additionally, because LCG does not photobleach, the enzymatic reaction can be monitored at several different time points, providing a general kinetic window for reactivity. Another benefit is that this assay can be run in the absence of the sensor components. At some time point an aliquot can be removed from the enzymatic reaction and added to the plate, followed by addition of the receptor/dye pairs. This would remove that possibility of competitive receptor inhibition or require the addition of other detection cocktails that could interfere with the enzyme activity.

The true utility of this approach for enzymatic assays is the ability to screen the reaction conditions in the presence of modified substrates. In this regard we can take advantage of the long-range perturbations of receptor binding to better understand how neighboring modifications can influence enzymatic activity. To test this, we ran the same mock kinase assay using a modified 'substrate', H3 1-12 K9$(Me)_3$, and performed a similar competitive experiment.

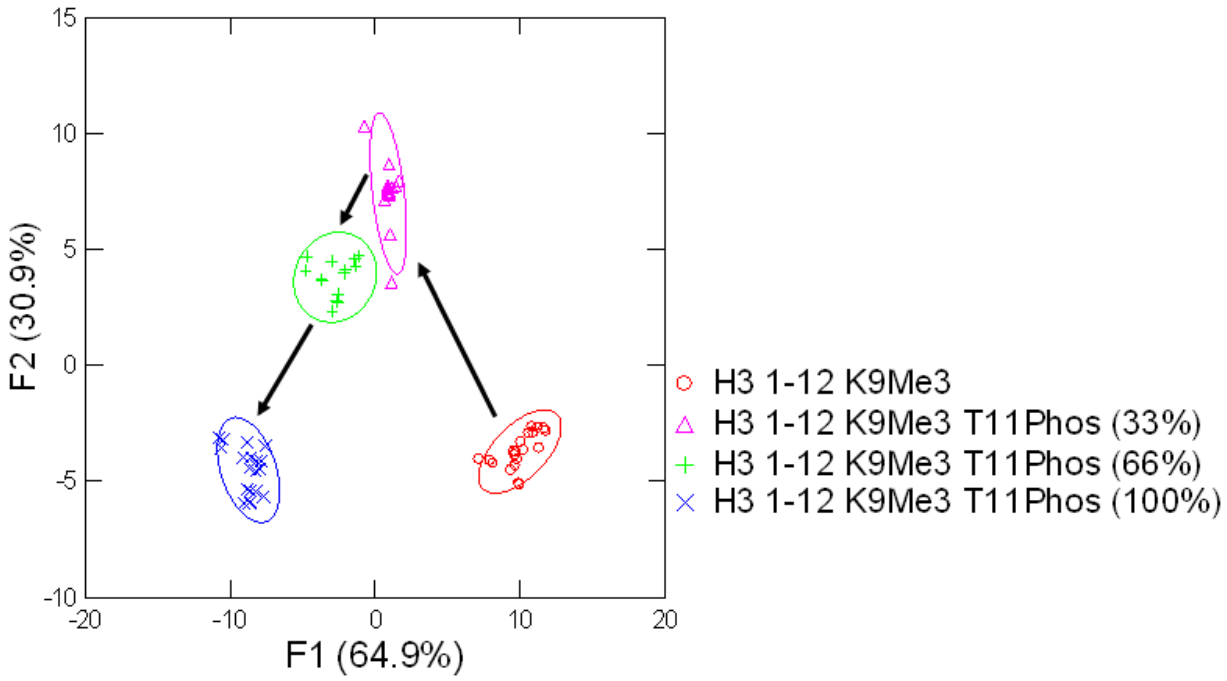


Figure 3.24 LDA of the mock enzymatic kinase reaction monitoring conversion of H3 1-12 K9$(Me)_3$ to H3 1-12 K9$(Me)_3$ T11 Phos. Arrows were added to represent the path of phosphorylation, confidence ellipses at 90%.

The resulting LDA plot for the mock kinase activity on the modified substrate again showed 100% classification and jackknife accuracy. While this approach would require a training set for each peptide substrate analyzed, the advantage is in not having to do a full assay redesign for each enzyme and substrate. We envision this assay could be utilized for a large number of histone modifying enzymes, including acetyltransferases, arginine deiminases, and methyltransferases and demethylases. The sensor array is applicable to each enzyme because they will modify a residue either by adding a mark responsible for binding to the receptor, as in methylation, or by reducing the affinity by removal of positive charge, as in acetylation. Each of these small changes should be recognized by the sensor array, allowing positive classification and qualitative monitoring of enzymatic reactions.

3.5 Experimental

3.5.1 Peptide Synthesis

All peptide synthesis was performed on a Tetras Peptide Synthesizer using CLEAR-Amide resin from Peptides International using Fmoc N-terminal protected amino acids with protected side chain functionality. Coupling reagents were HOBt/HBTU in DMF with 8 equivalents of DIPEA. After synthesis, all peptides were acylated using 5% acetic anhydride and 6% 2,6-lutidine in DMF, followed by cleavage and global deprotection using 95% TFA, 2.5% TIPS, and 2.5% H_2O for four hours. Trimethylated peptides were synthesized by adding 2 equivalents of Fmoc-Lys$(Me)_2$-OH. Following acylation, but prior to cleavage, the dimethyl lysine was further methylated using ten equivalents of MTBD and MeI. Peptides were purified using semi-preparative reverse phase HPLC using a XBridge Peptide C18 column with a linear gradient of A and B (Solvent A: 95% H_2O/5% CH_3CN with 0.1% TFA; Solvent B: 95% CH_3CN/5% H_2O with 0.1% TFA) and monitored at 214nm. Peptides were then lyophilized and characterized by ESI-MS.

3.5.2 Receptor Fluorescence Titration

Fluorescence quenching experiments were performed using purified receptors and commercially purchased Acridine Orange and Lucigenin. The titration used 5 μM of fluorophore and increasing concentrations of receptor in pH 10 50 mM glycine buffer. Receptor concentrations were determined using reported extinction coefficients.[4–6] Plates were centrifuged and incubated for 15 minutes prior to reading on a POLARStar Omega (BMG Labtech) using excitation: 485 nm, emission: 520 nm for AO and excitation: 370 nm, emission: 510 nm for LCG.

3.5.3 96-well Sensor Array

The 96 well sensor array utilized three sensors, **A_2B/LCG** (10 μM/2.5 μM), **A_2D/LCG** (5 μM/1 μM), and **A_2N** (15 μM/2.5 μM). The assay was run at 30 μL total volume per well at six replicates per analyte. Each well contained 10 μL of each peptide (final concentration 30 μM), receptor, and LCG. The plates were centrifuged for 1 minute and incubated for 15 minutes, followed by reading on a POLARStar Omega (BMG Labtech) using excitation: 370 nm, emission: 510 nm.

3.5.4 384-well Sensor Array

The 384 well sensor array utilized three sensors, **A_2B/LCG** (10 μM/2.5 μM), **A_2D/LCG** (5 μM/1 μM), and **A_2N** (15 μM/2.5 μM). The assay was run at 9 μL total volume per well at twenty replicates per analyte. Each well contained 3 μL of each peptide (final concentration 15 μM), receptor, and LCG. The plates were centrifuged for 1 minute and incubated for 15 minutes, followed by reading on a POLARStar Omega (BMG Labtech) using excitation: 370 nm, emission: 510 nm. Each analyte was run using a single plate, with control wells containing LCG alone and receptor/LCG to monitor the maximum fluorescence signal as well as the quenched state of the system.

3.5.5 Mock Enzyme Sensor Array

The mock enzyme assays were performed in 384 well plates, at volumes and concentrations of receptor/dye listed above. In addition to the plates run at 100% of either substrate or product peptide, two plates were analyzed at intermediate concentrations. Peptide solutions were pre-mixed in an eppendorf to make a cocktail at the appropriate concentration and

ratio such that when added to the well they were at 5:10 µM substrate:product for the 33% time point or 10:5 µM substrate:product for the 66% time point.

3.5.6 Linear Discriminate Analysis

LDA was performed using SysStat13. Prior to analysis, each analyte was normalized to the maximum fluorescence of the plates control well of LCG alone (F/F_{∞}). Additionally, each series of replicates was sorted from smallest response to largest response for each sensor, to decouple experimental error bias from the statistical analysis. The discriminate analysis was performed with all groups equal at 0.001 tolerance, and the resulting factors were plotted. The jackknife analysis was performed automatically in the software by classifying the data set while leaving one replicate at a time out, then re-submitting said replicate as a blind point to if the classification was upheld.

3.5.7 Sensor Array Statistical Validation

Sensor array validation was performed using Statistica Academic. The total replicate data for peptide samples was imported following the pre-processing described in the LDA. In excel, each replicate was randomly assigned a value of 1 or 2, to create two sample sets distributed at approximately 50% occupancy. The sample set of 1 was assigned as the training set, with sample set 2 as the test set. After import into statistica, the data was analyzed using general discriminate analysis with analyte labels as the dependent variable and the four sensors as the continuous predictors. The analysis was run using estimated probabilities of classifications using cross validation, with the analysis sample set to the training set 1. The analysis was run and the cross-validation percentage was saved. This analysis was repeated 20 times, randomizing the members

of each set prior to analysis. The average and deviation were then calculated to give a percentage of correct analysis for a 50/50 training test set.

REFERENCES

(1) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41.

(2) Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D. *Curr. Biol.* **2011**, *21*, 53.

(3) Rothbart, S. B.; Lin, S.; Britton, L.-M.; Krajewski, K.; Keogh, M.-C.; Garcia, B. a.; Strahl, B. D. *Sci. Rep.* **2012**, *2*, 1.

(4) Ingerman, L.; Cuellar, M. E.; Waters, M. L. *Chem. Commun.* **2010**, *46*, 1839.

(5) James, L. I.; Beaver, J. E.; Rice, N. W.; Waters, M. L. *J. Am. Chem. Soc.* **2013**, *135*, 6450.

(6) Pinkin, N. K.; Waters, M. L. *Org. Biomol. Chem.* **2014**, *12*, 7059.

(7) Beaver, J.; Peacor, B.; Bain, J.; James, L.; Waters, M. *Org. Biomol. Chem.* **2015**, *13*, 3220.

(8) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.

(9) Fischle, W.; Tseng, B. S.; Dormann, H. L.; Ueberheide, B. M.; Garcia, B. a; Shabanowitz, J.; Hunt, D. F.; Funabiki, H.; Allis, C. D. *Nature* **2005**, *438*, 1116.

(10) Kim, J. Y.; Banerjee, T.; Vinckevicius, A.; Luo, Q.; Parker, J. B.; Baker, M. R.; Radhakrishnan, I.; Wei, J. J.; Barish, G. D.; Chakravarti, D. *Mol. Cell* **2014**, *54*, 613.

(11) Rothbart, S. B.; Dickson, B. M.; Ong, M. S.; Krajewski, K.; Houliston, S.; Kireev, D. B.; Arrowsmith, C. H.; Strahl, B. D. *Genes Dev.* **2013**, *27*, 1288.

(12) Ramón-Maiques, S.; Kuo, A. J.; Carney, D.; Matthews, A. G. W.; Oettinger, M. a; Gozani, O.; Yang, W. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18993.

(13) Furey, T. S. *Nat. Rev. Genet.* **2012**, *13*, 840.

(14) Mann, M.; Jensen, O. N. *Nat. Biotech* **2003**, *21*, 255.

(15) Britton, L.; Gonzales-Cope, M.; Zee, B. M.; Garcia, B. *Expert Rev. Proteomics* **2011**, *8*, 631.

(16) Lothrop, A. P.; Torres, M. P.; Fuchs, S. M. *FEBS Lett.* **2013**, *587*, 1247.

(17) Rothbart, S. B.; Krajewski, K.; Strahl, B. D.; Fuchs, S. M. *Peptide microarrays to interrogate the "histone code,"* 1st ed.; Elsevier Inc., 2012; Vol. 512.

(18) Egelhofer, T. a; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. a; Cheung, M.-S.; Day, D. S.; Gadel, S.; Gorchakov, A. a; Gu, T.; Kharchenko, P. V; Kuan, S.; Latorre, I.; Linder-Basso, D.; Luu, Y.; Ngo, Q.; Perry, M.; Rechtsteiner, A.; Riddle, N. C.; Schwartz, Y. B.; Shanower, G. a; Vielle, A.; Ahringer, J.; Elgin, S. C. R.; Kuroda, M. I.; Pirrotta, V.; Ren, B.; Strome, S.; Park, P. J.; Karpen, G. H.; Hawkins, R. D.; Lieb, J. D. *Nat. Struct. Mol. Biol.* **2011**, *18*, 91.

(19) Huang, H.; Lin, S.; Garcia, B.; Zhao, Y. *Chem. Rev.* **2015**, *115*, 2376.

(20) Ong, S.-E.; Mann, M. *Nat. Protoc.* **2006**, *1*, 2650.

(21) Zhao, Y.; Jensen, O. N. *Proteomics* **2009**, *9*, 4632.

(22) Lavigne, J. J.; Anslyn, E. V. *Angew. Chemie Int. Ed.* **2001**, *40*, 3118.

(23) Umali, A. P.; LeBoeuf, S. E.; Newberry, R. W.; Kim, S.; Tran, L.; Rome, W. a.; Tian, T.; Taing, D.; Hong, J.; Kwan, M.; Heymann, H.; Anslyn, E. V. *Chem. Sci.* **2011**, *2*, 439.

(24) Minami, T.; Esipenko, N. a; Zhang, B.; Isaacs, L.; Anzenbacher, P. *Chem. Commun.* **2014**, *50*, 61.

(25) Liu, Y.; Minami, T.; Nishiyabu, R.; Wang, Z.; Anzenbacher, P. *J. Am. Chem. Soc.* **2013**, *135*, 7705.

(26) Stewart, S.; Ivy, M. A.; Anslyn, E. V. *Chem. Soc. Rev.* **2014**, *43*, 70.

(27) Friedman, J. H. *J. Am. Stat. Assoc.* **1989**, *84*, 165.

(28) Minaker, S. a.; Daze, K. D.; Ma, M. C. F.; Hof, F. *J. Am. Chem. Soc.* **2012**, *134*, 11674.

CHAPTER 4 Fingerprint Sensor Array for Neutral Histone Modifications

4.1 Background and Motivation

4.1.1 Lysine Acetylation

One of the most prevalent and studied modifications is the acetylation of lysine.[1,2] Acetylation is promoted by a family of enzymes known as the histone acetyl transferases (HATs) in combination with the cofactor acetyl CoA. Equally important however is the erasing of acetylation marks, which are catalyzed by the histone deactylases (HDACs), and which are proven drug targets, especially due to the large number of acetylated/deactylated sites in the histone tails.[3] Because a large driving force of chromatin compaction is the electrostatic interaction of the histone proteins and DNA, higher levels of acetylation have been known to weaken the compaction through loss of positive charge, leading to transcriptional activation.[4] In 2002, it was observed by the Burlingame group through mass spectrometry that the H4 tail is host to a dimethyl lysine residue at K20 that is responsible for attracting a HAT enzyme which acetylates the remaining lysine residues C-terminal to N-terminal in a zipper fashion, suggesting a mechanism for dissociating DNA from the histone protein.[5]

The HAT family of enzymes catalyzes the transfer of an acetyl group from the acetyl-CoA cofactor (Figure 4.1). This process utilizes a proton transfer through a water channel inside the active site prior to nucleophilic attack of the ε-amino group onto the acetyl carbonyl.[6]

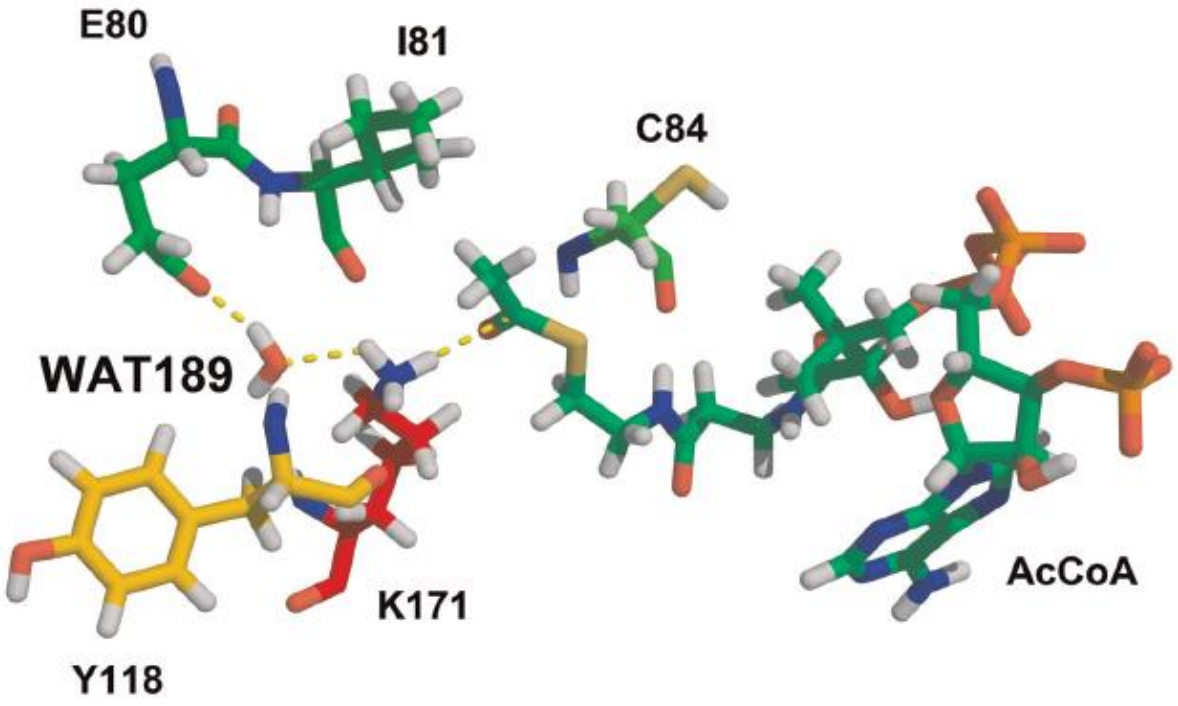


Figure 4.1 QM/MM calculation of the critical residues in the GCN5 HAT enzyme catalyzing the acetylation of K171. H-bonds responsible for deprotonation and acetylation shown in yellow.[6]

Radioactive labeling has been used to great success to study these enzymes using ^{14}C-AcCoA.[7] However, these assays suffer from the same pitfalls as other radiometric methods, including complex reaction work-up, end-point analysis, high cost, and danger of handling radioactivity.

4.1.2 Lysine Deacetylation

Because of the critical nature of histone acetylation in maintaining cell homeostasis, it is a highly regulated marker, with well-established enzymatic pathways for deacetylation. These eraser enzymes in the HDAC family are responsible for catalyzing the acetyl-lysine hydrolysis followed by release of acetate, shown in Figure 4.2.[8]

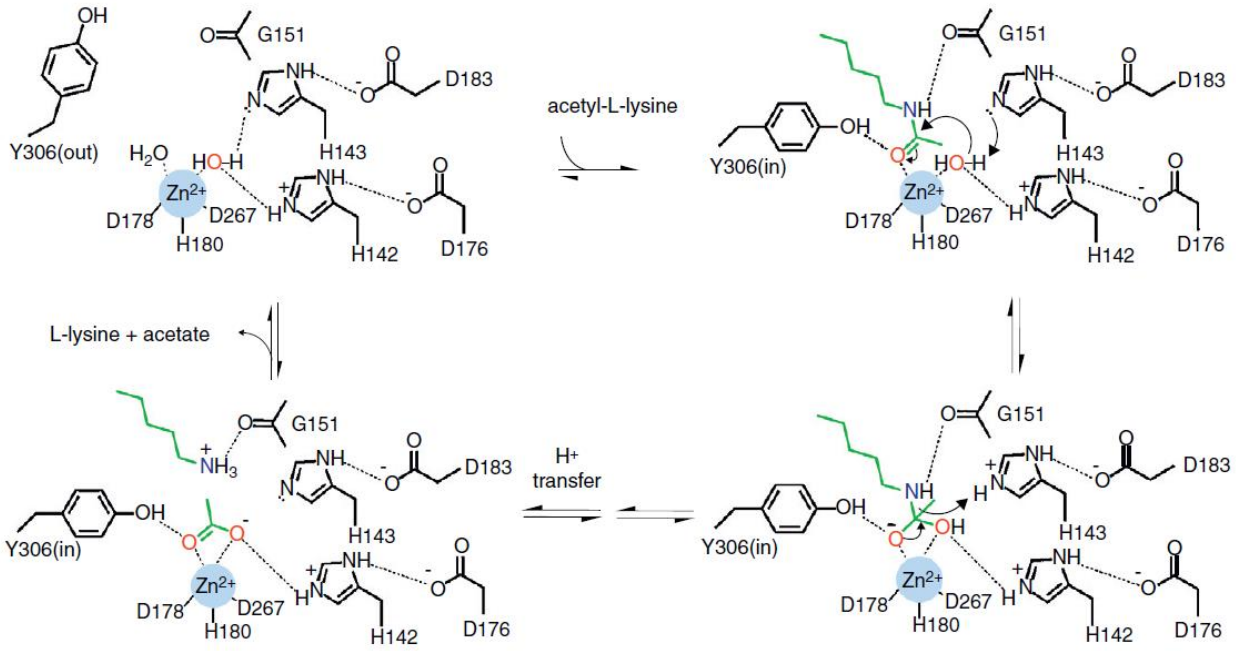


Figure 4.2 Proposed mechanism of the metal-dependent HDAC enzymes. The transition metal complexed in combination with neighboring histidine acts as a general base for the hydrolysis reaction.[8] Reprinted from *Current Opinion in Chemical Biology*, *21*, Lombardi, P.; Cole K.; Dowling, D.; Christianson D. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes, 735, copyright 2011, with permision from Elsevier.

The HDAC family of enzymes promotes transcriptional repression by allowing the reformation of condensed chromatin structure, and its misregulation has been implicated in a large variety of diseases.[9,10] This is such a prevalent marker in cancer that two HDAC inhibitors have been developed and are approved by the FDA for human use, vorinostat (Merck 2006) and romidepsin (Gloucester Pharmaceutical 2009). HDAC analysis is commonly done through MS or antibody methods, though fluorescence assays have also been developed. One such assay, described by the Fierke group in 2014, utilized a sequence of reactions to evaluate HDAC activity.[11] Time points from the reaction can be quenched and the acetate that formed up to that point is subjected to acetyl-CoA synthetase to reproduce acetyl-CoA, which can be further converted into citrate and CoA. This enzymatic conversion relies on the production of oxaloacetate, which is produced through a NAD^+ facilitated reaction, allowing the assay to

monitor the production of one mole of NADH per mole of acetate from the reaction. This requires a large number of moving parts to be optimized and operate correctly, in addition to not being sensitive to lower levels of acetate production, demonstrating that there is still a need for a high-throughput assay for monitoring acetylation and deacetylation.

4.1.3 Arginine Deimination

The deamination of arginine to produce the amino acid citrulline has been implicated in a number of cancers and diseases, including Parkinson's disease and Alzheimer's.[12] Citrullination abolishes the positive charge of arginine, and is proposed to act as an arginine methylation regulatory pathway by preventing methylation.[1] Additionally, citrullination enzymes have been shown to catalyze the demethylimination of monomethylargine, the only known method for removing an arginine methylation, though it does not furnish an unmethylated arginine species.[13]

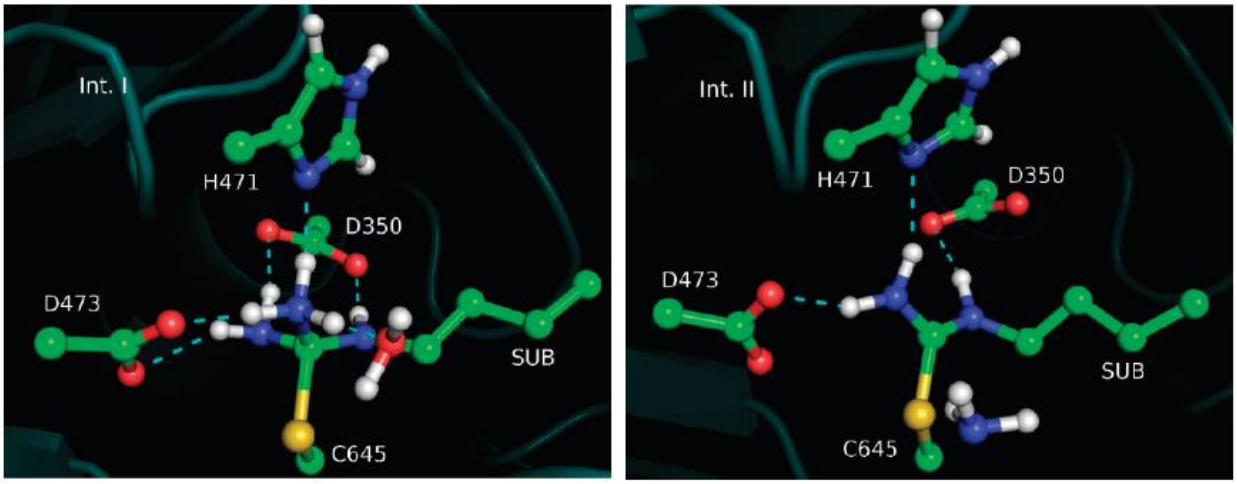


Figure 4.3 Molecular modeling of the PAD4 deimination reaction intermediates.[14] Reprinted (adapted) with permission from *J. Phys. Chem. B*, **2009**, *113*, 12750. Copyright 2009 American Chemical Society.

The deamination reaction is catalyzed by the PAD family of enzymes. The reaction proceeds through a tetrahedral intermediate (Figure 4.3 Int. I) formed by initial attack on the

guanidino carbon by an active site cysteine. This allows the elimination of ammonia from arginine (Figure 4.3 Int. II), followed by addition of a water molecule and subsequent elimination of cysteine to furnish the citrulline modified peptide.[15]

There are several methods of detecting arginine deimination beyond the use of antibodies and mass spectrometry.[16] A fluorescence HPLC method was used to monitor the depletion of substrate and growth of product, though are not well suited to complex peptides even though they show good specificity.[17] Another method developed in 2013 by the Lawrence lab utilized fluorescence, wherein a fluorescent PAD4 substrate is bound to a quenching dye through interaction with the positively charged arginine.[18] Upon deimination, the quencher is unbound, recovering fluorescence signal. This assay was expanded to monitor PAD4 activity in cell lysates as well as run simultaneously with a protein kinase assay, allowing the monitoring of multiple enzymatic pathways in tandem.

4.1.4 Motivation

The clinical importance of acetylation and citrullination events in the context of histones prompts the continual design and discovery of small molecule inhibitors and drug candidates for disease treatment. Coupled to these programs is the requirement of furthering the understanding of the HAT/HDAC and PAD family of enzymes capable of installing these modifications. However, the current chemical biology tools have issues in the detection and classification of the modifications, even in an endpoint assay. In 2012, the Strahl lab described the current state of acetylation antibodies, one of the standard methods of detection, as inadequate.[19] In a panel of site-specific acetyl antibodies, which passed all standard specificity testing, they found the preferred epitope was actually poly-acetylated substrates. This, coupled with other studies has shown the need for more robust analytical methods of classifying histone analytes.[20]

We have previously described the use of an indicator displacement assay (IDA) for detection and analysis of modified histone proteins (see Chapter 3). The IDA utilizes the fluorescence signal of dye displacement from small molecule receptors upon addition of a competitive analyte.[21,22] The receptors used, designed in the Waters lab, were characterized and studied using methylated lysine or arginine peptides, and are selective, moderate to high affinity binders for the higher degrees of methylation.[23–26] However, the binding events were also influenced by the neighboring sequence, namely the presence or absence of an adjacent basic residue, as shown in Table 4.1, in which Lys or Arg residues were mutated to the neutral Gly.[25]

Table 4.1 ITC Binding data for **A_2B** and **A_2N** to the Histone 3 (H3) peptide Ac-WGGG-QTA(R8X)(K9Z)STG-NH_2.[25]

Receptor	Peptide	K_d (μM)
A_2B	H3 K9	22 ± 1
A_2B	H3 R8G-K9	~140
A_2B	H3 R8-K9G	~ 150
A_2N	H3 K9	10.5 ± 0.9
A_2N	H3 R8G-K9	~360
A_2N	H3 R8-K9G	~300

These binding affinities showed that the receptor affinity was affected not just by loss of positive charge, but also by the identity of that charge. This implied that a sensor array approach could be useful in studying these acetylation and citrullination events. A sensor array relies on the general binding of substrates rather than specific interactions. Therefore, we hypothesize that each receptor will respond differently to the varying degrees of charge neutralization, caused by small difference in binding affinities. By combining the fluorescence readout of three general receptors, we could classify peptides bearing patterns of neutralized charge, potentially applying the system to complex biologically relevant events as well as establishing the potential for use in enzymatic assays.

4.2 Results

4.2.1 Single Sensor Citrulline Displacement

The work presented herein towards the discovery and optimization of neutral peptide sensors was performed with the assistance of Christopher Ramsay, an undergraduate in the Waters lab. Our initial experiments were designed to test if a sensor, comprised of the methyl lysine receptor **A_2B** and LCG, would be sensitive to the removal of positive charge. We chose to study the histone H3 peptide comprising of the first 1-12 amino acids, with a C-terminal tyrosine, because this peptide had two of each basic residue, lysine and arginine. This allowed us to study each type of neutral modification as well as the sequence dependence to determine if the neighboring landscape would influence binding and subsequent fluorophore displacement. Our proof of concept was to determine if the sensor **A_2B** would be responsive to the neutralization of one or both arginines, and if it could distinguish where the neutralization was taking place in the peptide sequence.

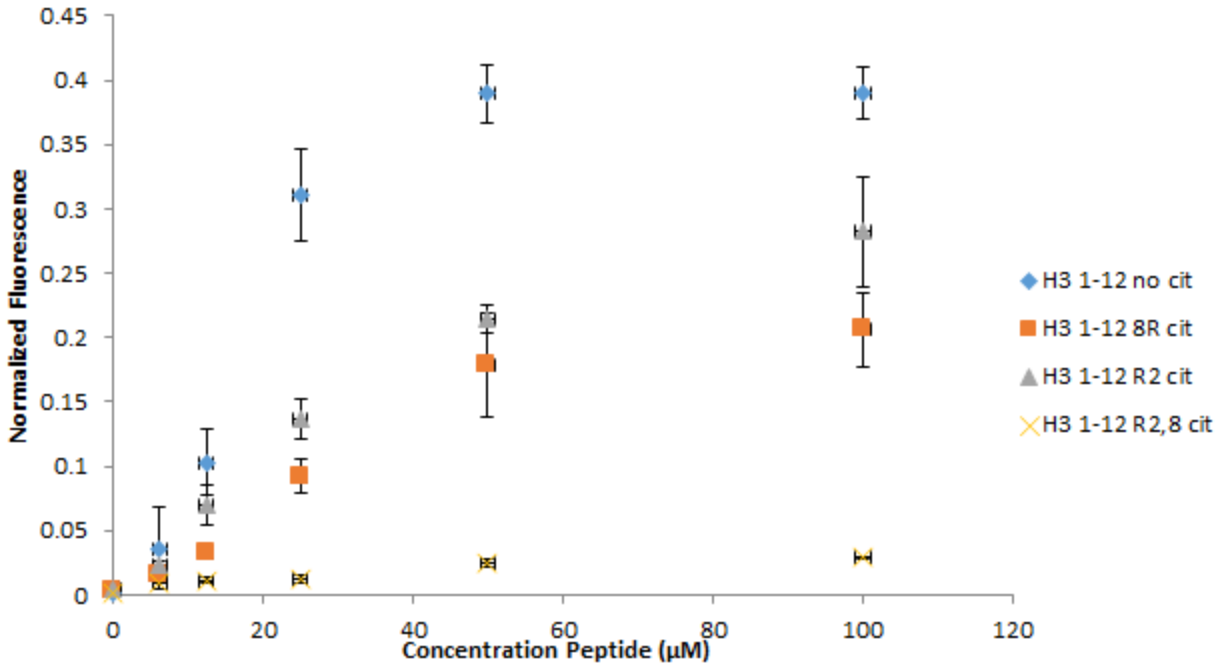


Figure 4.4 IDA for four different H3 peptides (Ac-ARTKQTARKSTGY-NH_2) with **A_2B** (10 μM) and LCG (2.5 μM). The fluorescence was normalized so that the quenched state (no peptide) was 0 while the fluorophore in solution was 1. Experiments run in triplicate in 50 mM glycine buffer, pH 9.15.

With these four peptides, we observed the expected trends in fluorescence, namely that decreasing the positive charge in the peptide lowered its binding affinity to the receptor. With one arginine neutralized it dropped to approximately half fluorescence, and two modifications resulted in almost no LCG displacement. However, as we anticipated, the single sensor has difficulty distinguishing the sequence of citrullination. Because we hypothesized that the receptor would still be able to interact with unmethylated lysine, the two sites of possible citrullination should have slightly different affinities, being either adjacent or spaced one residue away. While it seemed that the peptide signals were different, the fluorescence signals for the R2Cit and R8Cit peptides were close to within error of each other, suggesting that a single receptor is not sufficient.

4.2.2 Citrulline Sensor Array

Having demonstrated the applicability of a combined sensor array to classify peptides bearing identical modifications in different sequences, we believed that sensing the position of citrullination should be possible using three receptors, **A_2B**, **A_2D**, and **A_2N**. We performed a high-throughput 384 well plate experiment to determine if the array was able to correctly classify each peptide species modified with one citrulline at either arginine, or two citrulline modifications.

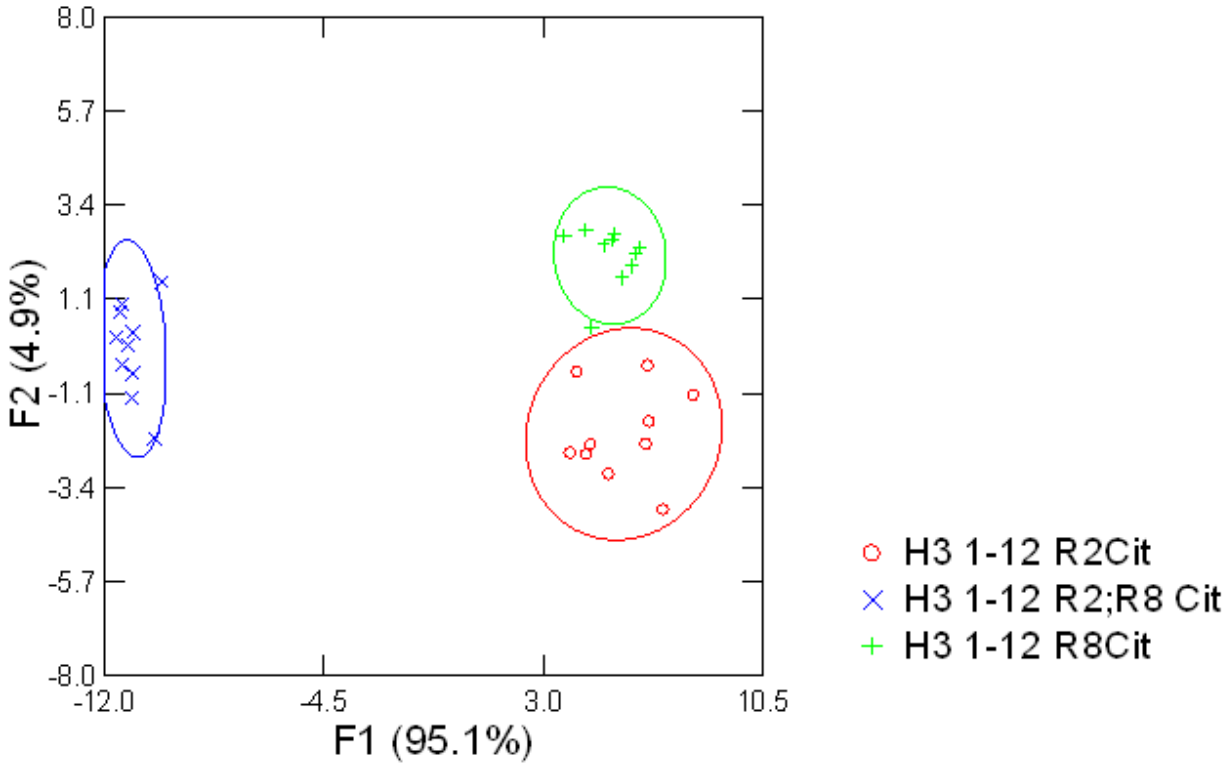


Figure 4.5 LDA plot of the fluorescence sensor array with the H3 peptide Ac-AXTKQTAXKSTGY-NH_2 (15 μM) where X is either Arg or Cit. **c** 50 mM glycine buffer, pH 9.15, confidence ellipses at 85%.

Gratifyingly, all three sensors were able to 100% classify a series of ten replicates for each peptide, notably separating the two mono-citrullination events from one another. This signifies that even with the minor differences in affinity these peptides are expected to have with

the receptors, it is enough to separate them into classes. To further examine this discriminatory capability, we wanted to determine a limit of detection for the analytes. In Figure 4.4 we observed that the signal decreases significantly at lower concentrations as the peptide affinities are reduced due to the loss of charge, with less than 10% recovery for the citrullinated peptides at 12.5 μM under the conditions measured. However, by applying all three sensors, even with low fluorescence signal we should be able to discriminate the three peptide species. We performed two experiments, lowering the concentration of the peptides while keeping the receptor and dye concentrations constant to determine an effective working range for detection.

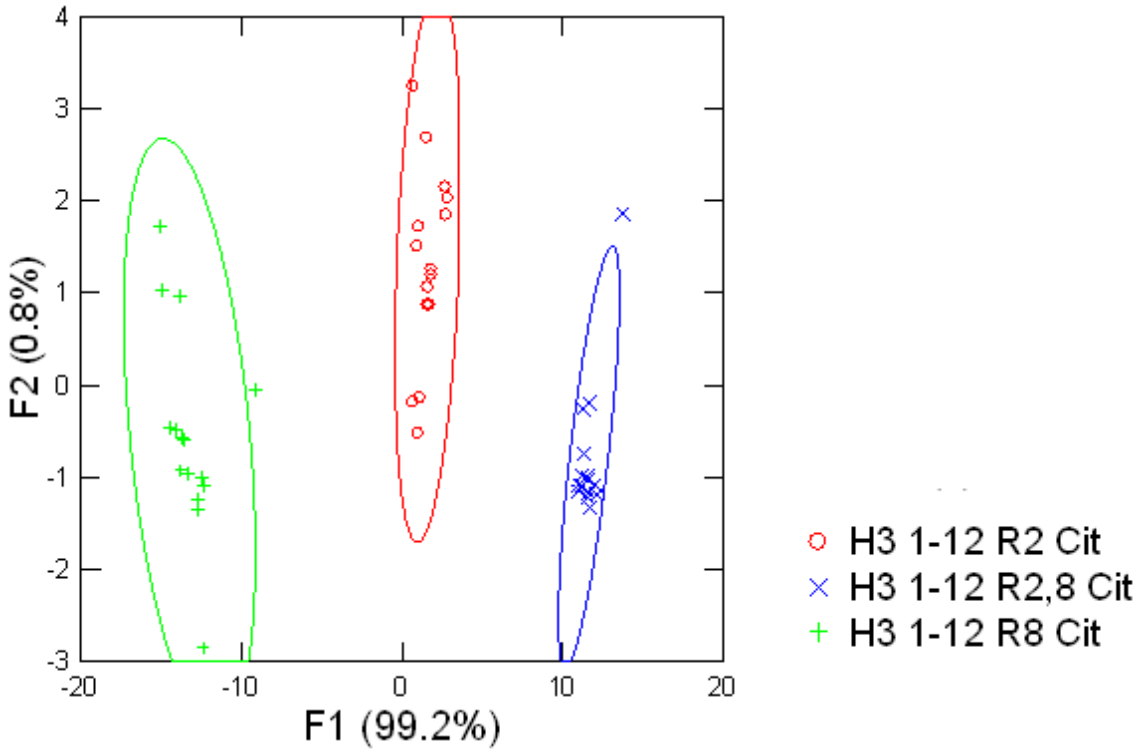


Figure 4.6 LDA plot of the fluorescence sensor array with the H3 peptide Ac-AXTKQTAXKSTGY-NH_2 (10 μM) where X is either Arg or Cit. **A_2B**/LCG (10 μM/1 μM), **A_2D**/LCG (5 μM/1 μM), and **A_2N** (15 μM/1 μM) in 50 mM glycine buffer, pH 9.15, confidence ellipses at 95%.

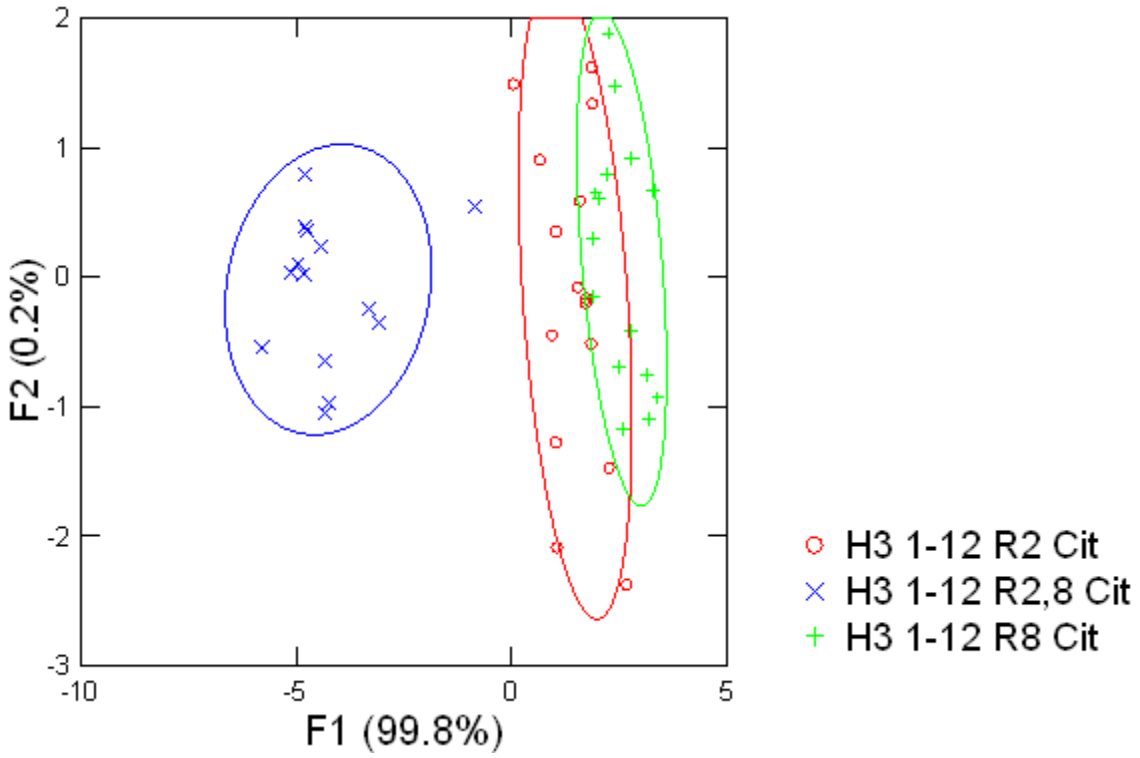


Figure 4.7 LDA plot of the fluorescence sensor array with the H3 peptide Ac-AXTKQTAXKSTGY-NH_2 (5 µM) where X is either Arg or Cit. **A_2B**/LCG (10 µM/1 µM), **A_2D**/LCG (5 µM/1 µM), and **A_2N** (15 µM/1 µM) in 50 mM glycine buffer, pH 9.15, confidence ellipses at 80%.

The sensor array platform is still operational at 10 µM of peptide, with 100% classification accuracy. This result aligns with the observed fluorescence in the single sensor titration (Figure 4.4), which still had discriminatory power at this range. However, lowering the peptide concentration to 5 µM significantly reduced the accuracy of the assay, with 84% initial classification and 78% accuracy in the jackknife analysis. Most of this inaccuracy comes from the mono-citrullinated peptides, since at the lower concentration of peptide the weaker affinity doesn't generate a signal strong enough to give discrimination of the two analytes.

4.2.3 Lysine Acetylation Sensor Array

In addition to arginine citrullination, we hypothesized that lysine acetylation could be easily studied using an identical platform. Previous fluorescence studies using **A_2N** (Figure 2.26) showed that by shortening the sequence of H3 from the first fifteen residues to the first twelve residues subtracted a lysine residue and lowered the resulting fluorescence displacement of the system. By that logic, acetylation should provide the same response, masking the positive charge and lowering the overall affinity. We studied the same sequence as the citrullination peptides, which contained two lysine residues that could be host to acetylation. These peptides were studied using the sensor array instead of doing a single sensor titration because we hypothesized that sensing the position of a single acetylation would be similarly difficult.

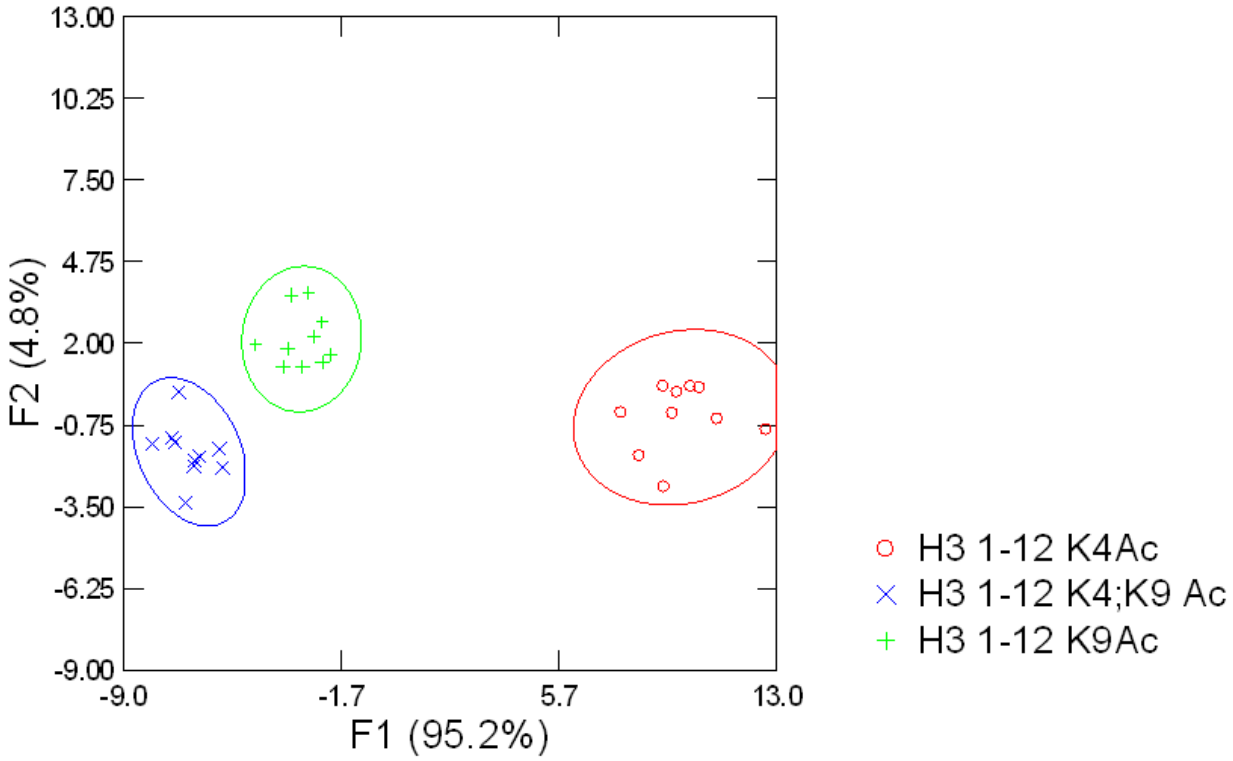


Figure 4.8 LDA plot of the fluorescence sensor array with the H3 peptide Ac-ARTXQTARXSTGY-NH_2 (15 µM) where X is either K or K(Ac). **A_2B**/LCG (10 µM/1 µM), **A_2D**/LCG (5 µM/1 µM), and **A_2N** (15 µM/1 µM) in 50 mM glycine buffer, pH 9.15, confidence ellipses at 90%.

Upon analysis of the sensor array we were able to classify the acetylated peptides with 100% accuracy. Interestingly, a quick examination of the output pattern suggests that the K9Ac peptide was more similar at this concentration to the doubly acetylated K4,K9 peptide than to the other mono-acetylated K4Ac peptide, suggesting that acetylation at lysine 9 has a stronger effect on lowering the peptide affinity. Nonetheless, the array was capable of distinguishing degree of neutralization as well as the sequence context. Analogous to the citrulline study we performed limit of detection analysis to determine the lowest effective concentration the assay can accurately classify while holding the receptor and dye constant.

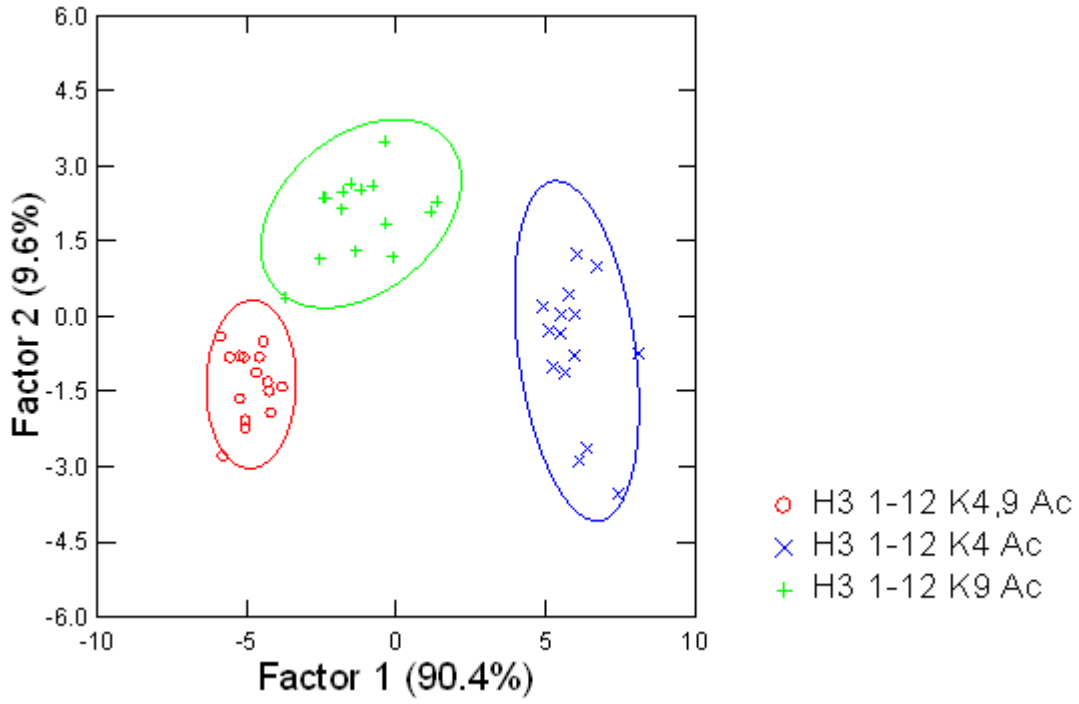


Figure 4.9 LDA plot of the fluorescence sensor array with the H3 peptide Ac-ARTXQTARXSTGY-NH_2 (10 µM) where X is either K or K(Ac). **A_2B**/LCG (10 µM/1 µM), **A_2D**/LCG (5 µM/1 µM), and **A_2N** (15 µM/1 µM) in 50 mM glycine buffer, pH 9.15, confidence ellipses at 90%.

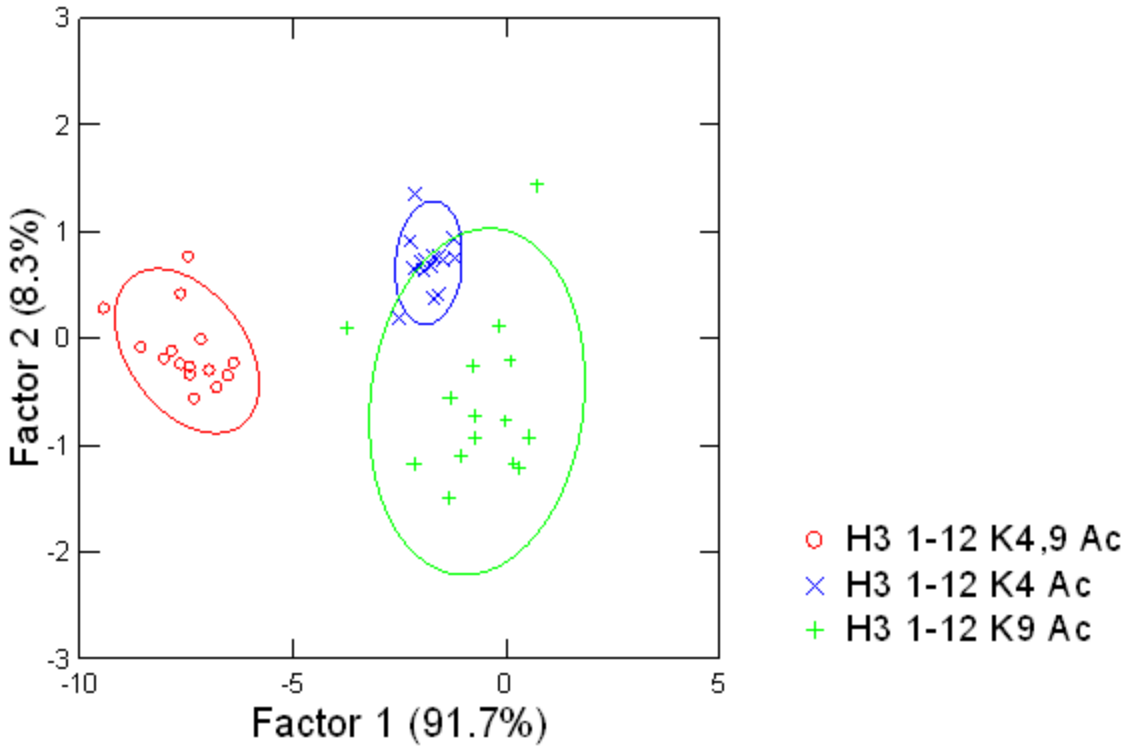


Figure 4.10 LDA plot of the fluorescence sensor array with the H3 peptide Ac-ARTXQTARXSTGY-NH_2 (5 µM) where X is either K or K(Ac). **A_2B**/LCG (10 µM/1 µM), **A_2D**/LCG (5 µM/1 µM), and **A_2N** (15 µM/1 µM) in 50 mM glycine buffer, pH 9.15, confidence ellipses at 90%.

As seen in the citrulline assay, the sensor was fully capable of discriminating each peptide species at 10 µM, with 98% classification accuracy, with the one misclass representing a K9Ac peptide classified as K4,9Ac. However, at 5 µM the accuracy decreased to 96%, this time arising from the overlap of the singly acetylated peptides as was observed for the citrulline experiment. Interestingly, this result indicates that the sensors are less sensitive to loss of lysine charge compared to arginine charge, suggesting that the arginine is more likely responsible for the fluorescence displacement of the unmodified H3 1-12 peptide, potentially due to more favorable electrostatic interactions with the carboxylate-decorated receptor.

4.2.4 Histone Neutralization Sensor Array

While the sensor array is able to distinguish between degrees of neutralization, it is not 100% clear that it is sensitive to the nature of the residue being neutralized, or if it is simply recognizing the loss of charge. Therefore, we studied the entire suite of peptides containing neutral PTMs described so far to determine if the array could correctly classify the degree of modification, the site of modification, and the identity of modification.

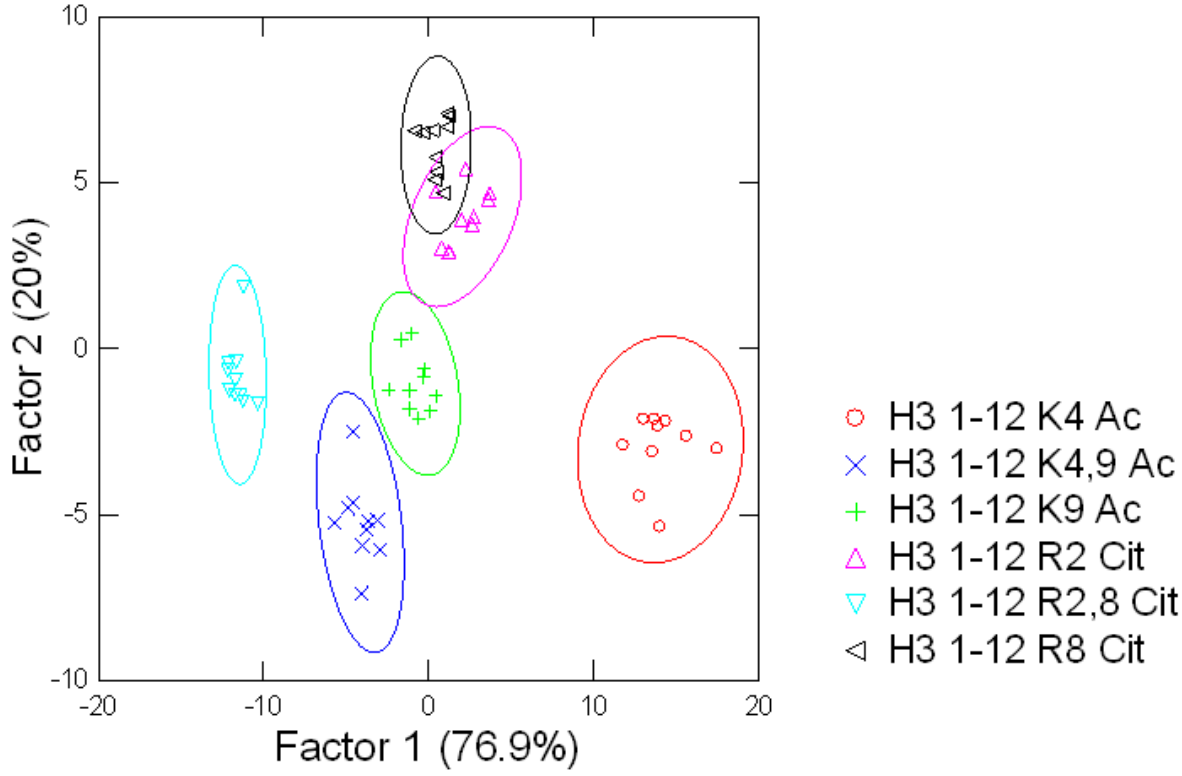


Figure 4.11 LDA plot of the combined sensor data for analytes bearing acetylation or citrullination (15 µM). **A$_2$B**/LCG (10 µM/1 µM), **A$_2$D**/LCG (5 µM/1 µM), and **A$_2$N** (15 µM/1 µM) in 50 mM glycine buffer, pH 9.15, 95% confidence ellipses.

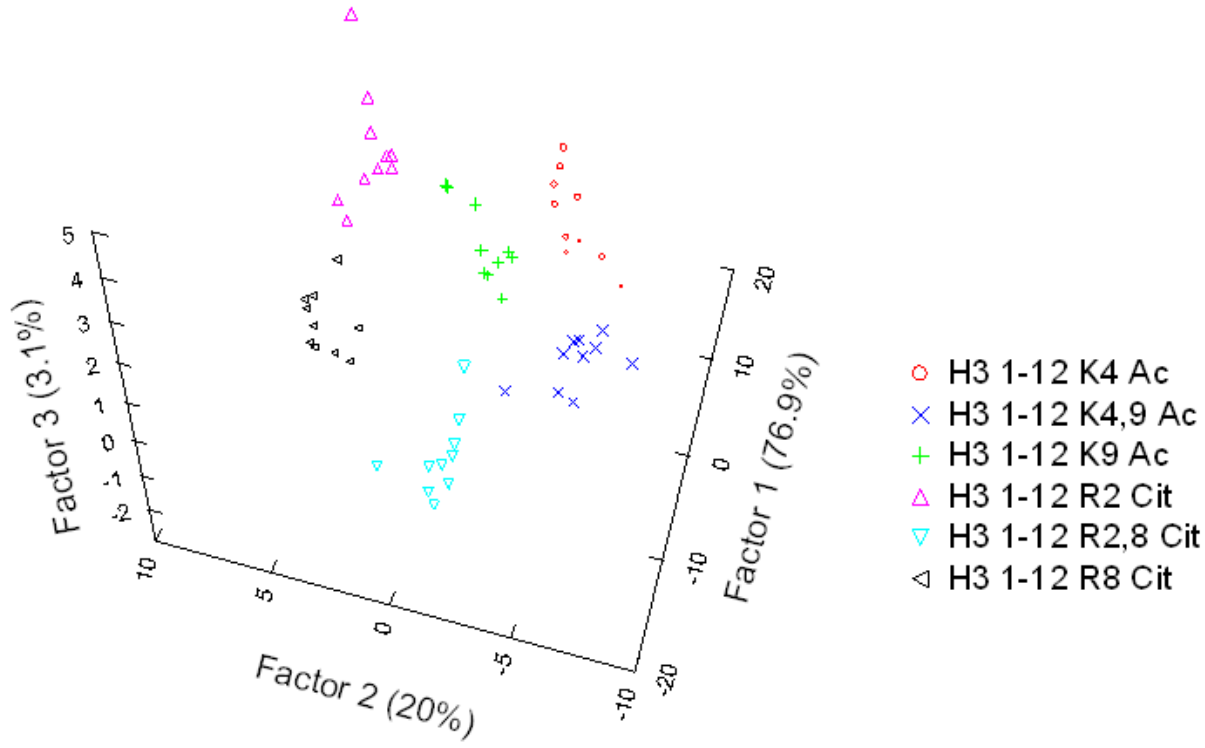


Figure 4.12 3D LDA plot of the combined sensor data for analytes bearing acetylation or citrullination (15 µM). **A$_2$B**/LCG (10 µM/1 µM), **A$_2$D**/LCG (5 µM/1 µM), and **A$_2$N** (15 µM/1 µM) in 50 mM glycine buffer, pH 9.15, 95% confidence ellipses.

The analysis of all six peptides bearing neutral modifications was able to successfully classify each analyte across ten replicates with 100% accuracy (Figure 4.12). This result shows that the array does not merely detect loss of charge, but rather differentiates between the subtle differences in binding affinity determined by at which residue and where in the peptide sequence the loss of charge was taking place. With the two factor output (Figure 4.11), there is some overlap in the single citrullination peptide signals, though plotting the third factor, responsible for 3% of the assay classification power solves this issue (Figure 1.12). These results confirm the sensor power of the small molecule receptor array, which will allow the further expansion into complex biological systems.

4.2.5 Future Directions

4.2.5.1 Poly Acetylation of Histone H4

Poly-acetylation is known to cause the dissociation of chromatin through a disruption of contacts across the nucleosome, by masking the positive charge required to associate with the anionic DNA.[4] One mechanism of acetylation proposed describes the acetylation of the histone tail in a zipper fashion, as shown in Figure 4.13.[5]

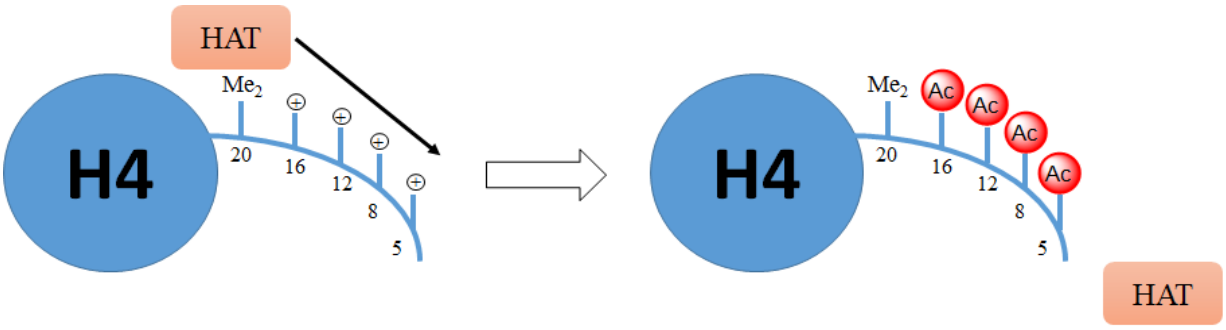


Figure 4.13 Zipper model describing histone H4 acetylation. Each lysine position is numbered starting from the N-terminus of the histone tail.

In this model, the HAT enzyme is first recruited to the K20$(Me)_2$, then moves down the histone installing acetylation events and opening the chromatin complex. The acetylation can be reversed, with the HDAC binding to the N-terminus K5 and proceeding in the reverse zipper format. This process is critical to activating transcription, but the poly-acetylated intermediates pose significant challenge for study. Currently there are antibodies capable of sensing a single acetylation or a pan acetylated H4 tail, but not the intermediate multiple acetylation events. An assay capable of distinguishing such complex histone landscapes would be highly advantageous, allowing the study of the intermediaries of this reaction.

With this goal in mind, we began work towards a sensor array platform that is capable of distinguishing between five different peptides, one for each major event in the H4 zipper acetylation. However, the H4 peptide poses a significant challenge, as it is highly positively

charged, with residues 15-20 consisting of a dense patch of basic residues, KRHRK. This could potentially complicate the analysis if a single acetylation is not able to sufficiently weaken the interaction of the analyte with our receptors. Additionally, our initial sensor array had minimal signal once we established a loss of two charges in the shorter peptide array. While this may not be the case starting from the longer peptide, it might be difficult to sense the final acetylation event as the peptide affinity might have weakened to such an extent that differentiation is impossible.

4.2.5.2 Addition of Sensors to the Array

With the challenges of sensing the lowest charge form of the H4 peptide as well as the highest charged species, we turned to several newly classified receptors for inclusion into the sensor array. The first of these, **A_2E**, was described previously, and was explored as a potential way to solve the issue of sensing weaker binders.[26] This receptor, with an extra carboxylate, was observed to bind to unmethylated lysine with the tightest affinity observed in the receptor family, at 6.7 μM. We screened the receptors ability to bind to LCG using a fluorescence titration, which can be seen in Figure 4.14.

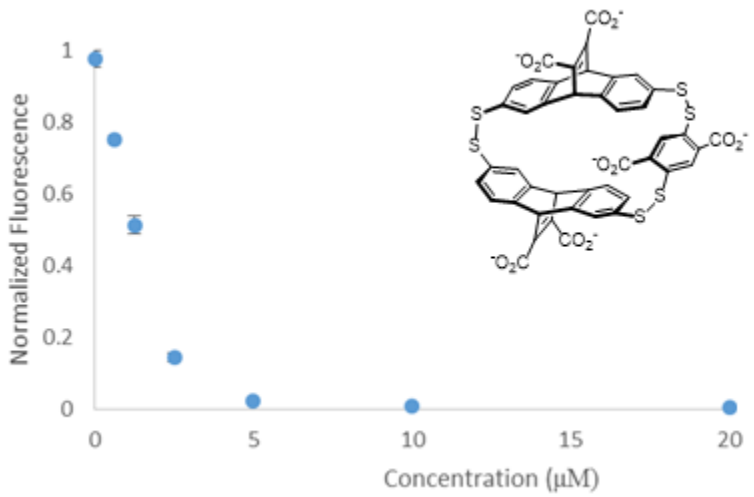


Figure 4.14 Fluorescence titration of **A$_2$E** (inset) into LCG (5 μM). Fluorescence was normalized for LCG alone (1) and the lowest observed fluorescence (0). 50 mM glycine buffer, pH 9.15.

The titration of **A$_2$E** and LCG revealed that the dye complexed quite strongly to the receptor, with a K_d of 0.7 μM. This makes the **A$_2$E** sensor the highest affinity for the fluorophore observed so far, which coupled with the tight affinity for the charged analytes should enable it to retain affinity as those charges are decreased. However, we believe this tight affinity will also significantly reduce the ability of **A$_2$E** to sense the single or even double acetylation events. Due to this, we wanted a weaker receptor that would potentially distinguish between unmodified and single acetylated H4, even if it had diminished sensitivity towards the multiply acetylated peptides.

Dr. Nick Pinkin, a former Waters lab member, discovered and characterized a number of receptors based on the **A$_2$B** and **A$_2$N** framework.[27] These receptors spaced the carboxylate functionality away from the binding pocket, as shown in Figure 4.15.

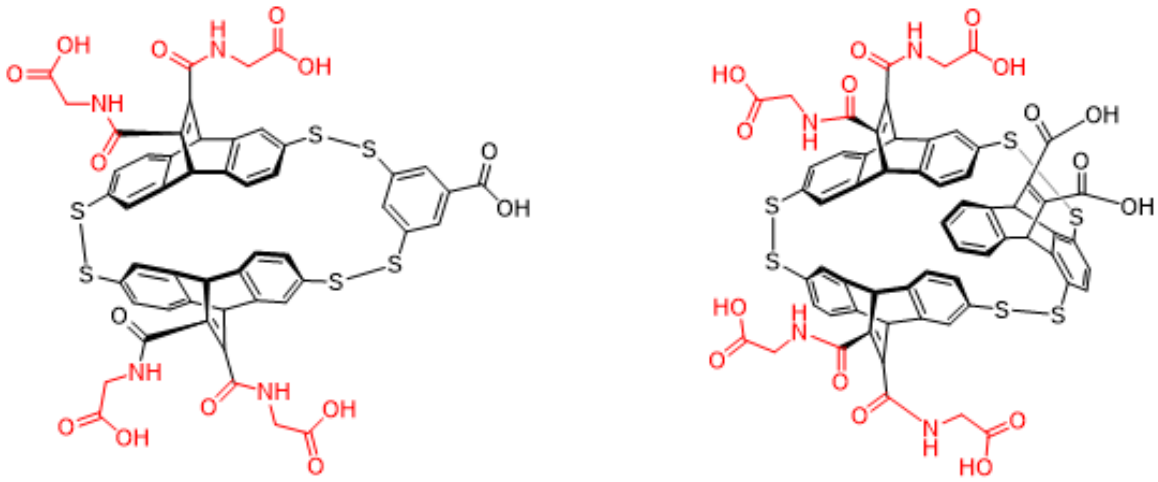


Figure 4.15 Carboxylate spaced receptors **GlyA$_2$B** (left) and **GlyA$_2$N** (right).

Table 4.2 ITC Data for **GlyA$_2$B** and **GlyA$_2$N** bound to Ac-WGGG-QTARK(Me)$_n$STG-NH$_2$.

	GlyA$_2$B		**GlyA$_2$N**	
	K_d (μM)	Selectivity	K_d (μM)	Selectivity
RK(Me)$_3$	7.6 ± 0.4	-	2.7 ± 0.7	-
RK(Me)$_2$	12.1 ± 0.1	1.6	5.5 ± 0.7	2
RK(Me)	30.8 ± 0.1	4.1	40 ± 3	14.7
RK	52 ± 3	6.8	60 ± 6	21.7
GK(Me)$_3$	99 ± 2	13	10.0 ± 0.6	3.6

The modified receptors still maintained selectivity for the methylated species of lysine, but the affinities dropped around two fold for the K(Me)$_2$ species, the hypothesized tightest binder in our sensor array for H4 zipper acetylation. Additionally, the affinities dropped between 3 and 6 fold for the unmethylated lysine, suggesting that the dense positive charge might have a lower affinity for these receptors. We proceeded to analyze both using fluorescence titrations to confirm LCG binding and quenching. Both glycine-spaced receptors quenched LCG, with affinities of 5.4 μM for **GlyA$_2$B** and 5.6 μM for **GlyA$_2$N**.

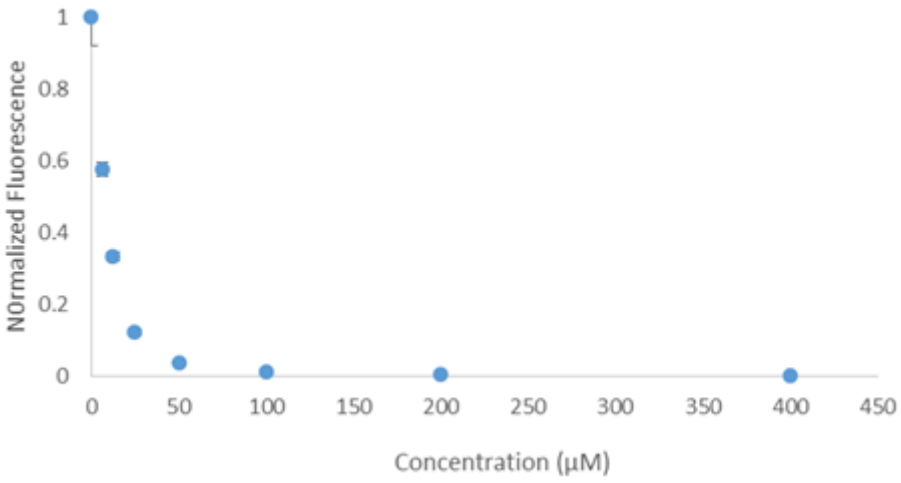


Figure 4.16 Fluorescence titration of **GlyA$_2$B** into LCG (5 µM). Fluorescence was normalized for LCG alone (1) and the lowest observed fluorescence (0). 50 mM glycine buffer, pH 9.15.

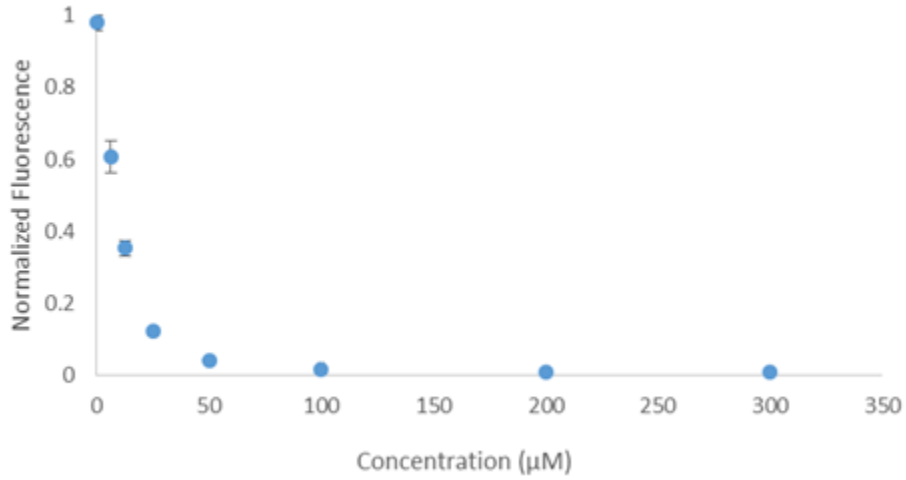


Figure 4.17 Fluorescence titration of **GlyA$_2$N** into LCG (5 µM). Fluorescence was normalized for LCG alone (1) and the lowest observed fluorescence (0). 50 mM glycine buffer, pH 9.15.

The addition of these glycine spaced receptors, as well as the tightly binding **A$_2$E**, should enable the study of the highly charged H4 peptide as well as the nearly neutralized polyacetylated H4 peptide, allowing classification of each unique acetylation point. Additionally, the sensor array could be easily adapted to an actual enzymatic assay using a process similar to that described in chapter 3, wherein a training set of varying peptides is established and

compared to time points of an enzymatic reaction. This sensor array, with its ability to classify both modified and unmodified peptides would enable the study of the HAT/HDAC family of enzymes as well as the PAD enzymes, which are traditionally much more challenging due to lack of easily monitored cofactor. This would provide a high throughput method for studying several critical factors, including histone code effect, enzyme mechanics, and can fuel inhibitor discovery, all based on the single, adaptable platform.

4.3 Experimental

4.3.1 Peptide Synthesis

All peptide synthesis was performed on a Tetras Peptide Synthesizer using CLEAR-Amide resin from Peptides International using Fmoc N-terminal protected amino acids with protected side chain functionality. Coupling reagents were HOBt/HBTU in DMF with 8 equivalents of DIPEA. After synthesis, all peptides were acylated using 5% acetic anhydride and 6% 2,6-lutidine in DMF, followed by cleavage and global deprotection using 95% TFA, 2.5% TIPS, and 2.5% H_2O for four hours. Peptides were purified using semi-preparative reverse phase HPLC using a XBridge Peptide C18 column with a linear gradient of A and B (Solvent A: 95% H_2O/5% CH_3CN with 0.1% TFA; Solvent B: 95% CH_3CN/5% H_2O with 0.1% TFA) and monitored at 214nm. Peptides were then lyophilized and characterized by ESI-MS.

4.3.2 Receptor Fluorescence Titration

Fluorescence quenching experiments were performed using purified receptors and commercially purchased Lucigenin. The titration used 5 μM of fluorophore and increasing concentrations of receptor in pH 10 50 mM glycine buffer. Receptor concentrations were determined using reported extinction coefficients.[23–25] Plates were centrifuged and incubated for

15 minutes prior to reading on a POLARStar Omega (BMG Labtech) using excitation: 370 nm, emission: 510 nm.

4.3.3 Peptide Fluorescence Displacement Experiments

Peptide titrations were performed in Costar 96 well half area black plates, NBS treated, or Corning 384 well half volume black NBS treated plates using **A_2B** (10 μM) and LCG (2.5 μM) described above. Each well contained the indicated concentrations of fluorophore and receptor with increasing concentrations of peptide. The fluorescence was normalized such that the highest fluorescence recovery over each of the experiments plotted was set to 1 while the lowest fluorescence signal was set to 0. Titrations were run in triplicate to ensure no major errors occurred. Peptide concentrations were determined using the extinction coefficient of Tyrosine, 1405 $cm^{-1}M^{-1}$ at 274 nm.

4.3.4 384-well Sensor Array

The 384 well sensor array utilized three sensors, **A_2B/LCG** (10 μM/1 μM), **A_2D/LCG** (5 μM/1 μM), and **A_2N** (15 μM/1 μM). The assay was run at 9 μL total volume per well at twenty replicates per analyte. Each well contained 3 μL of each peptide (final concentration 15 μM), receptor, and LCG. The plates were centrifuged for 1 minute and incubated for 15 minutes, followed by reading on a POLARStar Omega (BMG Labtech) using excitation: 370 nm, emission: 510 nm. Each analyte was run using a single plate, with control wells containing LCG alone and receptor/LCG to monitor the maximum fluorescence signal as well as the quenched state of the system.

4.3.5 Linear Discriminate Analysis

LDA was performed using SysStat13. Prior to analysis, each analyte was normalized to the maximum fluorescence of the plates control well of LCG alone (F/F_{∞}). Additionally, each series of replicates was sorted from smallest response to largest response for each sensor, to decouple experimental error bias from the statistical analysis. The discriminate analysis was performed with all groups equal at 0.001 tolerance, and the resulting factors were plotted.

REFERENCES

(1) Kouzarides, T. *Cell* **2007**, *128*, 693.

(2) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.

(3) Selvi, B. R.; Mohankrishna, D. V.; Ostwal, Y. B.; Kundu, T. K. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2010**, *1799*, 810.

(4) Kurdistani, S. K.; Tavazoie, S.; Grunstein, M. *Cell* **2004**, *117*, 721.

(5) Zhang, K.; Williams, K.; Huang, L.; Yau, P.; Siino, J.; Bradbury, E. M.; Jones, P.; Minch, M.; Burlingame, A. *Mol. Cell. Proteomics* **2002**, *1*, 500.

(6) Jiang, J.; Lu, J.; Lu, D.; Liang, Z.; Li, L.; Ouyang, S.; Kong, X.; Jiang, H.; Shen, B.; Luo, C. *PLoS One* **2012**, *7*, e36660.

(7) Dormeyer, W.; Ott, M.; Schnölzer, M. *Mol. Cell. Proteomics* **2005**, *4*, 1226.

(8) Lombardi, P. M.; Cole, K. E.; Dowling, D. P.; Christianson, D. W. *Curr. Opin. Struct. Biol.* **2011**, *21*, 735.

(9) Grunstein, M. *Nature* **1997**, *389*, 349.

(10) Falkenberg, K. J.; Johnstone, R. W. *Nat. Rev. Drug Discov.* **2014**, *2*, 673.

(11) Wolfson, N. a.; Pitcairn, C. A.; Sullivan, E. D.; Joseph, C. G.; Fierke, C. a. *Anal. Biochem.* **2014**, *456*, 61.

(12) Bicker, K. L.; Thompson, P. R. *Biopolymers* **2013**, *99*, 155.

(13) Wang, Y.; Wysocka, J.; Sayegh, J.; Lee, Y.-H.; Perlin, J. R.; Leonelli, L.; Sonbuchner, L. S.; McDonald, C. H.; Cook, R. G.; Dou, Y.; Roeder, R. G.; Clarke, S.; Stallcup, M. R.; Allis, C. D.; Coonrod, S. a. *Science* **2004**, *306*, 279.

(14) Ke, Z.; Zhou, Y.; Hu, P.; Wang, S.; Xie, D.; Zhang, Y. *J. Phys. Chem. B* **2009**, *113*, 12750.

(15) Leopoldini, M.; Marino, T.; Toscano, M. *Theor. Chem. Acc.* **2008**, *120*, 459.

(16) Hensen, S. M. M.; Pruijn, G. J. M. *Mol. Cell. Proteomics* **2014**, *13*, 388.

(17) Chikuma, T.; Yamada, M.; Tsuda, a; Yamamoto, M.; Nakashima, K.; Yajima, R.; Kato, T. *Anal. Biochem.* **2000**, *285*, 230.

(18) Wang, Q.; Priestman, M. a.; Lawrence, D. S. *Angew. Chemie - Int. Ed.* **2013**, *52*, 2323.

(19) Rothbart, S. B.; Lin, S.; Britton, L.-M.; Krajewski, K.; Keogh, M.-C.; Garcia, B.; Strahl, B. D. *Sci. Rep.* **2012**, *2*.

(20) Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D. *Curr. Biol.* **2011**, *21*, 53.

(21) Nguyen, B. T.; Anslyn, E. V. *Coord. Chem. Rev.* **2006**, *250*, 3118.

(22) Minaker, S. a.; Daze, K. D.; Ma, M. C. F.; Hof, F. *J. Am. Chem. Soc.* **2012**, *134*, 11674.

(23) Ingerman, L. a; Cuellar, M. E.; Waters, M. L. *Chem. Commun.* **2010**, *46*, 1839.

(24) James, L. I.; Beaver, J. E.; Rice, N. W.; Waters, M. L. *J. Am. Chem. Soc.* **2013**, *135*, 6450.

(25) Pinkin, N. K.; Waters, M. L. *Org. Biomol. Chem.* **2014**, *12*, 7059.

(26) E. Beaver, J.; C. Peacor, B.; V. Bain, J.; James, L. I.; L. Waters, M. *Org. Biomol. Chem.* **2015**, *13*, 3220.

(27) Pinkin, N. K. Discovery and Application of Small Molecule Receptors for the Recognition of Trimethyllysine in Water, University of North Carolina at Chapel Hill, **2015**.